

UNIVERSIDAD COMPLUTENSE DE MADRID.
FACULTAD DE ESTUDIOS ESTADÍSTICOS.
MÁSTER EN MINERÍA DE DATOS E INTELIGENCIA DE
NEGOCIO



USO Y ABUSO DE LAS TARJETAS BLACK POR LOS CONSEJEROS DE BANKIA

Trabajo Fin de Máster

Curso 2014/2015

Tutor del Proyecto: Conrado Miguel Manuel García

Autor del Proyecto: Yasmina Fernández Cuevas



INDICE

Contenido

1.	INTRODUCCION	3
1.1.	Planteamiento del problema.....	3
1.1.1.	Historia de Bankia.....	3
1.1.2.	Rescate del Gobierno a Bankia	4
1.1.3.	Tarjetas Black. Descripción de fenómeno	5
1.1.3.1.	¿Qué son las tarjetas Black?	5
1.1.3.2.	¿De dónde surge la denominación de tarjetas Black?	5
1.1.3.3.	Descubrimiento de las tarjetas.....	6
1.1.3.4.	Imputación de los consejeros.....	7
1.2.	Justificación de la necesidad de investigación	7
2.	OBJETIVOS	9
2.1.	Objetivo general	9
2.2.	Objetivos específicos.....	9
3.	METODOLOGÍA EMPLEADA.....	10
3.1.	Metodología de recogida de datos y variables de estudio	10
3.2.	Metodología de análisis de los datos	10
3.2.1.	Sample (Muestrear).....	10
3.2.2.	Explore (Explorar)	11
3.2.3.	Modify (Modificar)	11
3.2.4.	Model (Modelizar).....	12
3.2.4.1.	Análisis de correspondencias simples	12
3.2.4.2.	Análisis cluster	23
3.2.5.	Assesment (Evaluar)	30
3.3.	Software	30
4.	DESARROLLO DEL TRABAJO Y PRINCIPALES RESULTADOS	31
4.1.	Cronograma.....	31
4.1.1.	Fases o etapas de la investigación.....	31
4.1.2.	Periodo, lugar y materiales para el desarrollo de la investigación	31
4.1.3.	Cronograma	32
4.2.	Principales resultados.....	32
4.2.1.	Exploración de los datos. Resultados destacables del análisis descriptivo	32



4.2.2.	Resultados del análisis de correspondencias	42
4.2.3.	Resultados del análisis cluster jerárquico	45
4.2.4.	Resultados del análisis cluster no jerárquico (k-means)	47
4.2.5.	Resultados del análisis cluster difuso (fuzzy clustering).....	49
4.2.6.	Comparativa de resultados de los diferentes análisis cluster	51
5.	CONCLUSIONES	55
6.	BIBLIOGRAFIA	57

1.INTRODUCCION

1.1. Planteamiento del problema

1.1.1. Historia de Bankia

El **11 de septiembre de 1996** Miguel Blesa fue elegido como presidente de Caja Madrid, en sustitución de Jaime Terceiro. Fue reelegido en dos ocasiones, el **25 de octubre de 1999** y el **30 de septiembre de 2003**. Durante los años del mandato de Blesa (especialmente la última parte), la caja apostó por el crédito promotor.

El **22 de diciembre de 2008**, Esperanza Aguirre inició la batalla por el control de la entidad con la aprobación de una nueva Ley de Cajas que tenía por objetivo reducir la presencia de los representantes del Ayuntamiento de la capital, gobernado entonces por Alberto Ruiz Gallardón. Aguirre lucha para poner al frente de la caja al vicepresidente de la Comunidad de Madrid, **Ignacio González**, pero finalmente acepta a **Rodrigo Rato**, presionada por Mariano Rajoy.

La asamblea de la caja elige a Rodrigo Rato como presidente el **29 de enero de 2010** y se inicia la era de las macro fusiones. El **30 de julio** de ese mismo año, Caja Madrid anuncia el nacimiento de "la mayor caja de España" mediante su fusión con Bancaja (Caja de Ahorros de Valencia, Castellón y Alicante), Caja Insular de Ahorros de Canarias, Caja de Ahorros de la Rioja, Caja de Ahorros y Monte de Piedad de Ávila, Caja de Ahorros y Monte de Piedad de Segovia y Caixa d'Estalvis Laietana. Se constituye el Banco Financiero y de Ahorros (BFA). Un año más tarde, se alumbra la marca comercial del grupo, **Bankia**, y se solicita un préstamo de **4.465 millones** para afrontar los costes de la integración.

Bankia sale a **Bolsa** el **20 de julio de 2011** con un precio de 3,75 euros por acción, un 15% menos que el anunciado en el folleto entregado en la CNMV. El comportamiento en los primeros meses de cotización es estable, en contraposición al vivido por las bolsas mundiales, gracias a la intervención de las entidades cuidadoras, que se esforzaron por mantener en pie el valor.

Febrero de 2012 marca el declive final de la entidad. El día 2 de Febrero, el Gobierno anuncia una nueva reforma financiera, que obliga a las entidades a dotar de provisiones adicionales los activos inmobiliarios. Rodrigo Rato anuncia que Bankia puede hacerlo en solitario, pero las dudas crecen en torno a la entidad. El **10 de Febrero**, el valor inicia el rápido declive de su cotización en bolsa.

Mientras las diferencias entre Rodrigo Rato y el Gobierno de Rajoy van profundizándose, el FMI publica un informe en el que pide que Bankia mejore su balance y gestión. Poco después, la negativa de la auditora **Deloitte** de firmar las cuentas del ejercicio anterior precipitó el final de Rato y la intervención estatal.

Rodrigo Rato dimite el **7 de mayo de 2012** y propone a **José Ignacio Goirigolzarri** como sucesor.

El **9 de mayo**, BFA es nacionalizado y se convierte en el **mayor rescate bancario** de la historia del país. Goirigolzarri anuncia necesidades por 19.000 millones de euros, siendo finalmente el rescate de **más de 22.000 millones de euros**

1.1.2. Rescate del Gobierno a Bankia

El **7 de mayo de 2012**, el hasta entonces presidente de Bankia, Rodrigo Rato, presentó su renuncia proponiendo a José Ignacio Goirigolzarri como consejero y futuro presidente del grupo. La decisión coincidió con los planes del Gobierno de poner en marcha de forma inminente un plan de saneamiento para Bankia que preveía la inyección de cerca de 10.000 millones de dinero público con el objetivo de cubrir su riesgo inmobiliario.

El **9 de mayo de 2012**, se produjo la nacionalización del Banco Financiero y de Ahorros (BFA), matriz de Bankia. El Estado se quedó con todo el capital de BFA y, en consecuencia, se convirtió en dueño del 45% de Bankia. El Consejo de Bankia y de BFA nombró presidente a José Ignacio Goirigolzarri, en sustitución de Rato.

El **25 de mayo de 2012**, la CNMV suspendió la cotización de las acciones de Bankia a petición de la propia entidad. Por la tarde, se reunió el Consejo de Administración de Bankia que decidió pedir una inyección de 19.000 millones de euros de dinero público para BFA, (de los que 12.000 serían para Bankia), una cifra que desbordó todas las previsiones y que la convirtió en el mayor rescate financiero de la historia de España y uno de los mayores de Europa (y que sumada a los 4.465 millones ya concedidos, arrojaba un total de 23.465 millones de fondos públicos). A continuación el Consejo de Administración de Bankia dimitió en bloque.

El **28 de mayo de 2012**, las acciones de Bankia perdieron un 13,4% de su valor (habiendo perdido ya un 60% de su valor respecto al de salida en julio de 2011). El IBEX 35 cayó un 2,17 % (situándose en el nivel que tenía en mayo de 2003) y la prima de riesgo se disparó hasta alcanzar su máximo histórico desde la entrada en el euro, 511 puntos básicos.

A mediodía compareció en rueda de prensa el presidente del Gobierno **Mariano Rajoy** para anunciar que no "iba a haber ningún rescate de la banca española" y para descartar la exigencia de responsabilidades por la crisis de Bankia.

A última hora de la tarde del 28 de Mayo de 2012, el Consejo de Administración de BFA, anunció que tras reformular las cuentas de 2011, que en el momento de su publicación arrojaban beneficios, el banco tuvo unas pérdidas de 3.318 millones de euros, las mayores de la historia española (sólo por detrás de los 3.510 millones declarados por Banesto tras su intervención en 1993)

El crédito que inicialmente hizo el FROB a BFA por 4.465 millones de euros en el momento de su formación era imposible de pagar, por lo que el Estado, según el plan de emergencia propuesto cuando se formalizó el

crédito, convirtió ese crédito en acciones del FROB en BFA, y adicionalmente aportó otros 19.000 millones de euros más, el mayor rescate de una entidad financiera española, por lo cual el FROB, nacionalizó el 100% de BFA y se quedó como el único propietario de la entidad.

Al poseer entonces BFA el 45% de Bankia, el Estado se convirtió en el accionista mayoritario de esta entidad.

El **9 de junio de 2012**, el ministro de Economía **Luis de Guindos** anunció que España había solicitado y obtenido de la Unión Europea un rescate de hasta 100.000 millones de euros que el Estado utilizaría para sanear el sistema financiero español, especialmente Bankia, a través del FROB.

El presidente del Gobierno Mariano Rajoy en su comparecencia del mediodía del día siguiente evitó el término rescate y habló de "línea de crédito" y de éxito de su política.

El **27 de junio de 2012**, fuentes del Consejo de Administración del grupo de BFA-Bankia dieron a conocer la tasación realizada de cara a la nacionalización (-13.635 millones de euros). La consecuencia fue que los 4.465 millones aportados por el Estado, al convertirse de participaciones preferentes en acciones del BFA, se tradujeron en la nacionalización del 100% de la matriz e, indirectamente, del 45% de Bankia. Ante esta situación el Consejo de Administración de BFA dimitió en bloque.

1.1.3. Tarjetas Black. Descripción de fenómeno

1.1.3.1. ¿Qué son las tarjetas Black?

Se trata de unas tarjetas que se otorgaron a 86 miembros de la dirección y el Consejo de Administración de Caja Madrid y después de Bankia, destinadas a gastos de representación como comidas y viajes de negocios derivados del desempeño de las funciones de los consejeros. Bien utilizado se podría equiparar a una especie de bonus anual o compensación a su trabajo por un importe máximo de 50.000€.

Esta forma de remuneración es completamente legal si aparece en la Declaración de la Renta (tanto por parte de los beneficiarios como por parte de la empresa). Sin embargo quienes la recibían no tenían que justificar los gastos y hasta podían sacar dinero en efectivo de los cajeros.

Los gastos realizados con las mismas se camuflaban en cuentas que se utilizaban para **"errores del servidor informático"** creadas para hacer frente a "desajustes" de este tipo. Según el FROB la mayor parte de los 15,5 millones de euros gastados con estas tarjetas se asociaron durante 12 años a una cuenta llamada "Tratamiento Administrativo Circular 50/99", un concepto poco claro para descubrir el presunto fraude.

1.1.3.2. ¿De dónde surge la denominación de tarjetas Black?

Las tarjetas opacas de Bankia toman la denominación de "tarjetas Black" de dos vertientes fundamentales:

- Por una parte, de las "auténticas" tarjetas Black, que son productos financieros del más alto nivel, de los más exclusivos del mundo y apenas están dirigidos a un 1% de la población que la puede sufragar. Para poder acceder a estas tarjetas (cuyas comisiones de mantenimiento van desde los

\$95 hasta los \$1.500 anuales) hay que contar entre otros requisitos con unos ingresos muy elevados y dan a sus poseedores grandes beneficios como tratamiento VIP en viajes, entretenimiento..., reembolso en compras, diferentes tipos de seguros, horas de vuelo gratis en aviones de lujo...

De la similitud entre la exclusividad y grandes beneficios con las auténticas tarjetas Black, se toma el nombre para denominar a las tarjetas de Bankia

- Por otra parte, de la opacidad de las tarjetas ante Hacienda. La no declaración de los gastos realizados con estas tarjetas al fisco hacen que sean "Black" ante la Administración

1.1.3.3. Descubrimiento de las tarjetas

A diferencia de escándalos anteriores de la antigua caja, el caso de las tarjetas Black se destapó por decisión de los gestores de Bankia.

Un **e-mail** del 1 de Septiembre de 2009 con el asunto "Confidencia" del secretario saliente (De la Torre) al entrante (Rodrigo Rato), con copia a Miguel Blesa fue el detonante del descubrimiento del fraude de las tarjetas opacas de Caja Madrid, del cual se considera al ex-director de Caja Madrid, **Ildefonso Sánchez Barcoj**, como principal ideólogo.

Parte del e-mail decía lo siguiente: **"Los miembros de la Comisión de Control, conforme a la normativa de Caja Madrid, no pueden pertenecer a Consejos de filiales o participadas. Por tanto solo cobran dietas por las reuniones de la Comisión (1350 € brutos). Además, tiene cada uno una tarjeta visa de gastos de representación, Black a efectos fiscales hasta ahora (no está nada claro que la nueva jefa de inspección mantenga este criterio sobre todo teniendo en cuenta que Cipriano no conocía los nuevos importes), de 25000 € anuales excepto su Presidente que tiene una cobertura de 50.000 €"**.

El e-mail también muestra las dietas que cobraban los consejeros por pertenecer a distintos consejos, no solo los órganos de gobierno de la caja, sino también en las empresas participadas como Indra, Iberia, Mapfre, etc.

Cuando la información llega a oídos del Presidente, Goirigolzarri, encarga una auditoría para conocer la veracidad de los hechos.

La auditoría interna descubre esos 15,5 millones de euros gastados, que no se habían visto antes porque estaban escondidos bajo el epígrafe de "errores del servidor informático".

En este momento, Bankia contrata a unos asesores jurídicos que hacen un informe en el que, a grandes rasgos, se indica que la parte gastada durante la época de Bankia se puede reclamar, sin embargo la parte gastada durante la época de Caja Madrid no se puede reclamar.

En junio de 2012, Bankia pide a los cuatro directivos (Rodrigo Rato, José Manuel Fernández Norniella, Matías Amat e Ildefonso Sánchez Barcoj) que hicieron uso de las tarjetas que devuelvan lo que es del banco. Los cuatro directivos devuelven todo lo que gastaron (admitiendo implícitamente de esta forma la culpa)

Bankia eleva su informe al FROB para que sea éste quien busque la forma de perseguir los más de 15,2 millones de la época de Caja Madrid.

El FROB, accionista mayoritario, sólo pidió la devolución del dinero a quienes realizaron gastos entre 2011 y 2012. Posteriormente lo denunció ante la Fiscalía y le remitió toda la información.

1.1.3.4. Imputación de los consejeros

Desde **Octubre de 2014**, el juez de la Audiencia Nacional, Fernando Andreu, ha ido recabando información sobre la naturaleza de las tarjetas.

Además de la declaración de Rato, Blesa y Sánchez Barcoj, a los que impuso fianzas civiles de entre 3 y 16 millones de euros, el magistrado ha analizado las actas del consejo de administración de Caja Madrid.

Con fecha **28 de Enero de 2015**, Andreu imputa por apropiación indebida o administración desleal (la calificación definitiva todavía está por dilucidarse) a 78 miembros de los consejos de administración y directivos de Caja Madrid y de Bankia por usar para su beneficio personal las tarjetas de crédito opacas al fisco de Caja Madrid y Bankia entre 2003 y 2012. Estas imputaciones se suman a las realizadas en Octubre de 2014.

No obstante, en una primera fase cita a declarar solo a 27 imputados, todos ellos miembros de los consejos de administración. Las declaraciones se celebrarán en la Audiencia Nacional entre el 16 y el 18 de febrero de 2015.

La imputación de todos los consejeros y directivos había sido solicitada, entre otros por la Fiscalía Anticorrupción y la acusación popular que ejerce UPyD.

El magistrado ha incorporado también al sumario los correos electrónicos del ex-presidente de Caja Madrid Miguel Blesa referentes a las tarjetas opacas.

1.2. Justificación de la necesidad de investigación

La gravedad e interés de este tema radica no sólo en el posible uso fraudulento de las tarjetas por parte de los consejeros y directivos de la entidad, sino que el dato que agrava todavía más la situación es la realización de estos actos en una **entidad rescatada por el Estado con dinero público** de todos los ciudadanos, así como la actuación completamente insensible e irresponsable de algunos de los beneficiarios de estas tarjetas en los días e incluso horas previas al rescate de la entidad

Algunos de los datos más sorprendentes de los gastos realizados en fechas próximas al rescate son los siguientes:

- El **4 de Mayo de 2012**, mientras el ministro Luis de Guindos, en la sede de su Ministerio, reunió de urgencia a Rodrigo Rato, junto a los hombres más poderosos del país: Emilio Botín (Banco Santander), Isidro Fainé (La Caixa) y Francisco González (BBVA), para tratar el tema del rescate a España, Ildefonso Sánchez Barcoj hacía uso de su tarjeta Black en unos cines de Madrid.
- Ante la gravedad de la situación, Guindos volvió a citar a los jefes de la banca española el **6 de Mayo de 2012**.

Entre una reunión y otra, el miembro del Comité Ejecutivo José Manuel Fernández Norniella aprovechó el sábado para hacer la compra, y el día de la reunión definitiva utilizó su tarjeta Black para pagar la comida en un restaurante con el dinero de la entidad que estaba a punto de ser rescatada con dinero público.

- El fin de semana previo al rescate, el ex-vicepresidente del Gobierno se pasó por el cajero a retirar 1.000 euros el sábado, entre la primera y segunda reunión en el Ministerio de Economía.

A la noticia de la inyección de los primeros 10.000 millones de euros de dinero público, el lunes 7 de mayo, Rato reaccionó pagando una cena en un restaurante con su tarjeta Black.

- El director financiero, Ildefonso Sánchez Barcoj, “celebró” el rescate pasando por el cajero y sacando 300 euros.

2.OBJETIVOS

2.1. Objetivo general

El objetivo general de este Trabajo Fin de Máster (TFM) es realizar un estudio de los gastos realizados por los consejeros de la entidad con las tarjetas Black, así como buscar una posible agrupación de los consejeros por el uso que dio a la correspondiente tarjeta.

Para ello se empleará la información pública disponible con los cargos realizados con cada una de las tarjetas.

2.2. Objetivos específicos

Los objetivos específicos que inicialmente se plantean para la consecución del objetivo general anteriormente expuesto son los siguientes:

- Analizar el gasto general realizado con las tarjetas para obtener una visión global de dónde y cómo se gastaban los consejeros el dinero.
- Analizar el gasto realizado por cada uno de los consejeros con la intención de detectar comportamientos de gasto impropios de tarjetas de empresa supuestamente destinadas a gastos de representación.
- Segmentación de los consejeros para buscar patrones de gasto común.

3. METODOLOGÍA EMPLEADA

3.1. Metodología de recogida de datos y variables de estudio

Los **datos** relativos a los gastos realizados por cada uno de los consejeros se pueden obtener de la web de forma pública en varias fuentes. En este caso se han obtenido los datos del periódico “El País” en la dirección de su página web: <http://elpais.com/especiales/2014/tarjetas-opacas-caja-madrid/>.

Una vez descargados los datos de los gastos de los 82 consejeros y directivos de Bankia de la página web, se han unido en un mismo archivo, el cual cuenta con un total de **75.502 observaciones realizadas entre el 1 de Enero de 2003 y el 29 de Agosto de 2012**.

Las **variables** a considerar en nuestro estudio serán las siguientes:

- Nombre del comercio: variable cualitativa nominal que indica dónde se efectuó el cargo
- Tipo de gasto: variable cualitativa nominal que indica la actividad del comercio
- Importe: variable cuantitativa continua. Indica el importe en euros cargado en la tarjeta
- Operación: variable cualitativa nominal que indica el tipo de operación realizada
- Nombre del consejero: variable cualitativa nominal que se refiere al consejero o directivo que hizo uso de la tarjeta
- Fecha: variable cualitativa ordinal, con formato dd/mm/aaaa
- Hora: variable cualitativa ordinal, con formato hh:mm:ss

3.2. Metodología de análisis de los datos

La metodología a emplear para el análisis de los datos será la **metodología SEMMA** (Sample, Explore, Modify, Model, Assesment) en la que se llevarán a cabo las siguientes etapas:

3.2.1. Sample (Muestrear)

Es la primera etapa del proyecto. En ella se preparan los datos para su posterior exploración.

De esta forma, se han modificado las variables originales “Fecha” y “Hora”:

- *Fecha*: se ha sustituido esta variable por cuatro variables obtenidas a partir de esta como son:
 - Mes: variable cualitativa ordinal con valores según el mes en que se efectuara el cargo (Enero, Febrero,..., Diciembre)
 - Día del mes: variable cualitativa ordinal con valores enteros entre 1 y 31 según el día del mes en que se efectuara el cargo
 - Año: variable cualitativa ordinal con valores de 2003 a 2012 según el año en que se efectuara el cargo.
 - Día de la semana: variable cualitativa ordinal con valores según el día de la semana en que se efectuara el cargo (lunes, martes,..., domingo).

- *Hora*: se ha sustituido esta variable con datos de la hora, minutos y segundos en que se realizó el gasto por la variable “Hora” con únicamente la hora en que se hizo el gasto. Se trata de una variable cualitativa con valores enteros entre 0 y 23 según la hora del día en que se efectuara el cargo.

3.2.2. Explore (Explorar)

Se ha realizado la exploración de los datos tanto de forma gráfica como numérica.

- Exploración numérica: se ha realizado el estudio tanto del importe gastado como de la frecuencia de utilización de la tarjeta por cada una de las variables anteriormente indicadas (tipo de gasto, mes, día de la semana,...).

Esta exploración se ha realizado en dos vertientes:

- Por una parte, con el gasto total realizado en conjunto por todos los consejeros.
 - Por otra parte, estudiando el gasto y frecuencia de uso de la tarjeta para cada uno de los consejeros en cada una de las variables.
- Exploración gráfica: se ha realizado también la exploración gráfica de los datos y, del mismo modo que en el caso anterior, se ha realizado por una parte la exploración de los datos globales y por otra parte la exploración de los gastos correspondientes a cada uno de los consejeros individualmente.

Asimismo, se ha realizado un **análisis de correspondencias simples** entre la variable nombre del consejero y las distintas variables anteriormente mencionadas (tipo de operación, categoría de gasto, tramo horario,...)

Los resultados más destacables de estas exploraciones se detallan en el apartado 4.2 del presente documento.

3.2.3. Modify (Modificar)

Para realizar la segmentación de los datos, en primer lugar se ha categorizado la variable “Hora” en 3 categorías:

- **Mañana**: gastos realizados entre las 7:00 a.m. y las 14:00 p.m.
- **Tarde**: gastos realizados entre las 15:00 p.m. y las 22:00 p.m.
- **Noche**: gastos realizados entre las 23:00 p.m. y las 6:00 a.m.

Por otra parte, la variable “Tipo de Gasto”, que cuenta en un principio con 449 valores diferentes, se ha categorizado en las siguientes 11 categorías: **Cajeros y tarjetas; Gasolineras; Hoteles; Moda; Ocio; Pequeño Comercio; Restaurantes; Supermercados y Grandes Superficies; Telefonía y tecnología; Varios; Viajes.**

Para la segmentación de los consejeros, se ha generado un fichero con una única observación por cada consejero y las siguientes variables para cada uno de ellos:

- Importe y frecuencia de uso para las 11 categorías de “Tipo de Gasto” (gasolineras, hoteles, Moda,...)
- Importe y frecuencia de uso en cada uno de los años (desde 2003 hasta 2012)
- Importe y frecuencia de uso en cada uno de los meses (desde Enero hasta Diciembre)
- Importe y frecuencia de uso en cada uno de los días (desde el Lunes hasta el Domingo)
- Importe y frecuencia de uso en las 3 franjas horarias (mañana, tarde y noche)

Para tener los valores de las variables en una misma escala y poder obtener resultados coherentes en los análisis posteriores, necesitaremos **tipificar** estas variables. Para ello se resta la media y se divide por la desviación típica. Es decir, si tengo la variable formada por los datos x_1, x_2, \dots, x_n , entonces estos mismos datos tipificados z_1, z_2, \dots, z_n se definen como:

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

3.2.4. Model (Modelizar)

3.2.4.1. Análisis de correspondencias simples

El objetivo del Análisis de Correspondencias Simples es la visualización de Tablas de frecuencias de dos variables cualitativas, por lo que sirve para detectar las relaciones existentes entre dos variables nominales. En el presente trabajo se busca encontrar la relación existente (en el caso de que la haya) entre la variable nombre de consejero y el resto de las variables (tipo de gasto, categoría del gasto, día de la semana,...) para poder explicar el comportamiento de los consejeros.

Se detallan a continuación los análisis realizados en los que se ha podido detectar relación entre las variables.

ANÁLISIS CORRESPONDENCIAS SIMPLES ENTRE NOMBRE DE CONSEJERO Y CATEGORÍA DE GASTO

Lo primero que debo ver es la tabla de inercias y de la chi-cuadrado (Tabla 1)

Inercia y descomposición chi-cuadrado					
Valor singular	Inercia principal	Chi-cuadrado	Porcentaje	Porcentaje acumulado	9 18 27 36 45
0.71236	0.50745	5681182	46.56	46.56	*****
0.40209	0.16168	1810077	14.83	61.39	*****
0.35687	0.12735	1425775	11.68	73.08	*****
0.30400	0.09242	1034673	8.48	81.56	*****
0.24512	0.06008	672652	5.51	87.07	***
0.22319	0.04981	557698	4.57	91.64	***
0.17155	0.02943	329472	2.70	94.34	**
0.15735	0.02476	277176	2.27	96.61	*
0.14352	0.02060	230599	1.89	98.50	*
0.12770	0.01631	182574	1.50	100.00	*
Total	1.08989	1.22E7	100.00		
Grados de libertad = 760					

Tabla 1: Tabla de Inercias y Chi-Cuadrado

En esta tabla veo que con los dos primeros ejes tengo el 61,39% de la inercia recogida, que al ser mayor que el 60% lo puedo dar como válido.

A continuación veo si debo eliminar alguna variable

En primer lugar analizo los resultados FILA (Tabla 2)

Estadísticos de sumariación para los puntos de la fila				Estadísticos de sumariación para los puntos de la fila				Estadísticos de sumariación para los puntos de la fila				Estadísticos de sumariación para los puntos de la fila			
	Calidad	Mass	Inercia		Calidad	Mass	Inercia		Calidad	Mass	Inercia		Calidad	Mass	Inercia
36	0	0	0,0001	24	0,2918	0,0042	0,012	39	0,901	0,0112	0,0061	1	0,5517	0,0219	0,0104
62	0,0548	0,0007	0,009	43	0,332	0,0042	0,0051	46	0,9448	0,0114	0,0085	72	0,5681	0,022	0,0096
3	0,3161	0,0008	0,0009	74	0,3358	0,0042	0,0027	45	0,8551	0,0117	0,0243	80	0,5143	0,022	0,0156
57	0,3766	0,0015	0,0008	14	0,3079	0,0046	0,0052	58	0,1792	0,0118	0,0173	70	0,2323	0,0227	0,0223
18	0,8712	0,0016	0,0009	42	0,2151	0,0052	0,0051	35	0,4617	0,012	0,0121	11	0,2816	0,0234	0,0167
37	0,6395	0,0017	0,0013	4	0,8433	0,0055	0,0115	64	0,2844	0,0122	0,0083	48	0,6118	0,0243	0,0095
55	0,8532	0,0017	0,0021	16	0,5633	0,0056	0,0028	13	0,6116	0,0124	0,0067	22	0,4382	0,0248	0,0194
66	0,1297	0,0017	0,0021	25	0,2081	0,0061	0,0081	40	0,3326	0,0132	0,0079	27	0,3121	0,0254	0,0146
53	0,6529	0,0018	0,0034	33	0,2112	0,0061	0,0073	50	0,3825	0,0152	0,009	9	0,1367	0,0256	0,0222
68	0,8553	0,0018	0,0049	38	0,2737	0,0064	0,0088	19	0,2926	0,0155	0,0224	77	0,9843	0,0272	0,0692
23	0,1942	0,0019	0,0025	12	0,9407	0,0069	0,0064	29	0,3347	0,0155	0,0041	59	0,401	0,0316	0,0113
61	0,3204	0,0021	0,0008	21	0,5582	0,007	0,0036	20	0,5761	0,0157	0,0057	34	0,917	0,0348	0,0381
44	0,186	0,0031	0,0032	75	0,554	0,007	0,0041	32	0,1663	0,0165	0,0188	15	0,0725	0,0378	0,0095
79	0,3403	0,0032	0,0016	78	0,3217	0,0071	0,0107	65	0,2932	0,0169	0,0132	6	0,4472	0,0383	0,0322
69	0,3393	0,0034	0,0027	41	0,7974	0,0088	0,0064	28	0,7479	0,0171	0,0071	51	0,9922	0,0397	0,1301
30	0,4537	0,0035	0,0021	63	0,1759	0,0088	0,0028	67	0,4733	0,0183	0,0086	52	0,9965	0,04	0,1073
31	0,3846	0,0035	0,0046	49	0,2848	0,0089	0,0048	7	0,3362	0,0185	0,0126	73	0,5547	0,0511	0,0242
56	0,1728	0,0035	0,0066	2	0,3678	0,0095	0,0055	71	0,1375	0,0186	0,0367				
47	0,178	0,0039	0,005	17	0,3477	0,0095	0,0056	26	0,7839	0,0205	0,01				
5	0,3533	0,0042	0,0025	60	0,9514	0,0108	0,0192	76	0,6539	0,0207	0,0032				

Tabla 2: Estadísticos de sumariación para los puntos de la fila

En este caso vemos valores de la masa (**MASS**) muy pequeños respecto al máximo, por lo que miro si debo eliminar alguna fila. En el caso de que viéramos alguna medida muy pequeña deberíamos convertirla en suplementaria y obtener su proyección o unirla a otra fila.

En general podemos poner restricciones por debajo de 0.03 y ver cuál es el máximo. Si lo supera en 5 veces ($5 \times 0.03 = 0.15$) entonces debo realizar alguna de las dos opciones anteriores. Si esto no se cumple dejo la fila y continúo con el análisis.

En este caso, el máximo es de 0,0511. Como $0,0511 < 0,15$, no elimino ninguna fila.

En segundo lugar analizo los resultados COLUMNA (Tabla 3)

Estadísticas de sumariación para los puntos de columna			
	Calidad	Mass	Inercia
TELEF_TECNOL	0,0761	0,0104	0,0311
OCIO	0,1351	0,0276	0,052
VARIOS	0,1555	0,0276	0,032
GASOLINERAS	0,2578	0,0469	0,0535
MODA	0,2664	0,0537	0,0577
P_COMERCIO	0,2602	0,0728	0,0396
HOTELES	0,1433	0,073	0,0337
VIAJES	0,378	0,1002	0,0916
SUPERS	0,2035	0,1711	0,0977
CAJEROS_TARJETAS	0,9989	0,1772	0,3774
RESTAURANTES	0,9253	0,2395	0,1338

Tabla 3: Estadísticos de sumariación para los puntos de la columna

El valor máximo de estos datos es 0,2395 (correspondiente a “Restaurantes”) que es mayor a $5 \times 0.03 = 0.15$, por lo tanto incluiré como suplementarias todas las categorías con valor inferior a 0,03, por tanto, “Telef_tecnol”, “Ocio” y “Varios”.

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

Si volvemos a hacer el análisis con estas tres modalidades como suplementarias obtenemos los siguientes resultados:

Inercia y descomposición chi-cuadrado					
Valor singular	Inercia principal	Chi-cuadrado	Porcentaje	Porcentaje acumulado	10 20 30 40 50
0.71213	0.50713	5366544	50.79	50.79	*****
0.40248	0.16199	1714237	16.22	67.01	*****
0.36610	0.13403	1418283	13.42	80.43	*****
0.26008	0.06764	715792	6.77	87.20	***
0.25376	0.06439	681411	6.45	93.65	***
0.19499	0.03802	402343	3.81	97.46	**
0.15929	0.02537	268489	2.54	100.00	*
Total	0.99858	1.057E7	100.00		
Grados de libertad = 539					

Tabla 4: Tabla de Inercias y Chi-Cuadrado

En la Tabla 4 veo que con los dos primeros ejes tengo el 67,01% de la inercia recogida que al ser mayor que el 60% lo puedo dar como válido. Por lo tanto será suficiente con 2 dimensiones.

Análisis los resultados FILA (Tabla 5)

Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de fila				Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de fila				Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de fila				Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor		Dim1	Dim2	Mejor		Dim1	Dim2	Mejor		Dim1	Dim2	Mejor
2	0	0	1	47	0	0	1	1	2	0	2	39	0	0	2
5	0	0	1	50	0	0	1	3	0	0	2	41	0	2	2
7	1	0	1	51	1	0	1	4	0	2	2	43	0	0	2
11	1	0	1	52	1	1	1	6	0	2	2	45	0	2	2
13	0	0	1	54	0	0	1	9	0	0	2	46	0	2	2
14	0	0	1	56	0	0	1	12	0	2	2	48	0	2	2
16	0	0	1	57	0	0	1	15	0	0	2	49	0	0	2
17	0	0	1	58	0	0	1	18	0	0	2	53	0	0	2
20	0	0	1	61	0	0	1	19	0	2	2	55	0	0	2
21	0	0	1	64	0	0	1	22	0	2	2	59	0	2	2
23	0	0	1	65	1	0	1	24	0	2	2	60	0	2	2
25	0	0	1	66	0	0	1	26	2	2	2	62	0	0	2
27	1	0	1	67	0	0	1	28	2	0	2	63	0	0	2
30	0	0	1	70	1	0	1	29	0	0	2	68	0	2	2
32	0	0	1	71	1	0	1	31	0	0	2	69	0	0	2
33	0	0	1	73	1	0	1	35	0	2	2	72	0	2	2
34	1	0	1	74	0	0	1	36	0	0	2	75	0	0	2
40	0	0	1	77	1	0	1	37	0	0	2	76	0	0	2
42	0	0	1	79	0	0	1	38	0	0	2	78	0	2	2
44	0	0	1									80	0	2	2

Tabla 5: Índices de las coordenadas que mejora contribuyan a la inercia para los puntos de la fila

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila		
	Dim1	Dim2		Dim1	Dim2		Dim1	Dim2		Dim1	Dim2
59	0,0582	0,7903	49	0,1296	0,3073	29	0,2668	0,1004	64	0,2804	0,0124
60	0,2199	0,7238	35	0,1773	0,2939	73	0,4546	0,0961	51	0,9815	0,0122
68	0,1433	0,6629	26	0,5247	0,2909	25	0,3602	0,0726	11	0,2926	0,0122
45	0,1605	0,6483	37	0,4166	0,2548	54	0,3846	0,0667	32	0,2031	0,0113
12	0,3001	0,6447	48	0,4083	0,2545	52	0,9387	0,0593	65	0,3268	0,0096
4	0,1701	0,6237	31	0,2264	0,2518	57	0,4511	0,0561	66	0,2582	0,008
80	0,0098	0,6168	75	0,4096	0,246	14	0,2395	0,0556	20	0,6989	0,0076
55	0,2928	0,5947	19	0,0996	0,2407	30	0,432	0,0527	70	0,2476	0,0071
62	0,2013	0,5731	28	0,5953	0,2239	23	0,2791	0,0502	17	0,4426	0,007
76	0,2229	0,5708	43	0,1587	0,1944	2	0,365	0,0492	50	0,3758	0,0043
18	0,2758	0,5566	16	0,6187	0,1905	42	0,2291	0,0413	40	0,3343	0,0028
46	0,4411	0,5279	63	0,0046	0,1777	79	0,4259	0,0401	77	0,9812	0,0024
72	0,0233	0,5249	1	0,4241	0,1734	5	0,3356	0,0364	47	0,1974	0,0009
6	0,0482	0,4713	9	0,0359	0,1504	71	0,1713	0,0362	44	0,2174	0,0005
22	0,057	0,4566	38	0,1229	0,1457	56	0,773	0,0296	74	0,3676	0,0003
53	0,168	0,4316	69	0,365	0,1169	67	0,4839	0,0261	7	0,3541	0,0002
41	0,4208	0,4001	61	0,4656	0,1157	33	0,2101	0,0235	34	0,9166	0,0001
39	0,5753	0,3872	21	0,4945	0,1145	58	0,348	0,0211	36	0	0
24	0,0352	0,3116	3	0,201	0,1129	15	0,0087	0,0184			
78	0,0813	0,3098	13	0,5852	0,1107	27	0,3203	0,0157			

Tabla 6: Cosenos cuadrados para los puntos de la fila

Con la Tabla 5 y la Tabla 6 obtengo la siguiente explicación de los ejes:

- La dimensión 1 está más influenciada por los consejeros MORADO IGLESIAS, RICARDO (51) y TORRE MARTÍNEZ, ENRIQUE DE LA (77)
- La dimensión 2 está más influenciada por los consejeros PÉREZ CLAVER, MARIANO (59) y PÉREZ FERNÁNDEZ, FRANCISCO JOSÉ (60)

En segundo lugar analizo los resultados COLUMNA

Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de columna				Cosenos cuadrados para puntos de columnas		
	Dim1	Dim2	Mejor		Dim1	Dim2
CAJEROS_TARJETAS	1	0	1	CAJEROS_TARJETAS	0,9944	0,0049
GASOLINERAS	0	0	1	GASOLINERAS	0,2207	0,0422
HOTELES	0	0	2	HOTELES	0,0907	0,0965
MODA	0	2	2	MODA	0,0923	0,2079
P_COMERCIO	0	0	2	P_COMERCIO	0,0889	0,135
RESTAURANTES	2	2	2	RESTAURANTES	0,3536	0,548
SUPERS	0	0	2	SUPERS	0,1822	0,0673
VIAJES	0	2	2	VIAJES	0,0286	0,4232

Tabla 7: Resultados para los puntos de la columna

Con la Tabla 7 obtengo la siguiente explicación de los ejes:

- La dimensión 1 está más influenciada por el gasto en cajeros y tarjetas
- La dimensión 2 está más influenciada por el gasto en viajes

Los resultados de este análisis se detallan en el apartado 4.2. del presente documento.

ANÁLISIS CORRESPONDENCIAS SIMPLES ENTRE NOMBRE DE CONSEJERO Y OPERACION

Lo primero que debo ver es la tabla de inercias y de la chi-cuadrado (Tabla 8)

Inercia y descomposición chi-cuadrado					
Valor singular	Inercia principal	Chi-cuadrado	Porcentaje	Porcentaje acumulado	11 22 33 44 55
0.84222	0.70934	4273425	56.05	56.05	*****
0.69881	0.48833	2941952	38.59	94.64	*****
0.23082	0.05328	320967	4.21	98.85	**
0.10437	0.01089	65627	0.86	99.71	
0.06023	0.00363	21857	0.29	99.99	
0.00955	0.00009	550	0.01	100.00	
0.00037	0.00000	1	0.00	100.00	
0.00009	0.00000	0	0.00	100.00	
Total	1.26556	7624377	100.00		
Grados de libertad = 416					

Tabla 8: Tabla de Inercias y Chi-Cuadrado

En esta tabla (Tabla 8) veo que con el primer eje tengo el 56,05% de la inercia recogida, y con los dos primeros ejes ya obtengo el 94,64% de la inercia recogida, por lo que emplearé 2 ejes.

A continuación veo si debo eliminar alguna variable

En primer lugar analizo los resultados FILA (Tabla 9)

Estadísticos de sumarización para los puntos de la fila				Estadísticos de sumarización para los puntos de la fila				Estadísticos de sumarización para los puntos de la fila			
	Calidad	Mass	Inercia		Calidad	Mass	Inercia		Calidad	Mass	Inercia
36	0	0	0	43	0,9964	0,0078	0,0019	46	0,3797	0,0213	0,0135
62	0,9964	0,0014	0,0003	74	0,9964	0,0079	0,0019	58	0,9964	0,0219	0,0052
3	0,9964	0,0015	0,0004	5	0,9964	0,008	0,0019	13	0,9964	0,0231	0,0055
18	0,9496	0,0029	0,0013	24	0,9964	0,0081	0,0019	40	0,9964	0,0246	0,0059
57	0,9964	0,0029	0,0007	14	0,9964	0,0086	0,002	50	0,9964	0,0283	0,0067
37	0,9964	0,0031	0,0007	42	0,9964	0,0099	0,0024	19	0,9964	0,0289	0,0069
53	0,9964	0,0033	0,0008	4	0,9964	0,0102	0,0024	32	0,9964	0,0309	0,0074
55	0,9964	0,0033	0,0008	16	0,9964	0,0105	0,0025	72	0,9329	0,041	0,0039
66	0,9964	0,0033	0,0008	33	0,9964	0,0113	0,0027	1	0,9964	0,0421	0,01
68	0,9964	0,0033	0,0008	25	0,9964	0,0114	0,0027	70	0,9895	0,0425	0,01
23	0,9964	0,0038	0,0009	38	0,9032	0,012	0,0008	48	0,9964	0,0452	0,0108
61	0,9317	0,0039	0,0011	75	0,9964	0,013	0,0031	27	0,9955	0,0474	0,0103
44	0,9964	0,0059	0,0014	78	0,5261	0,0132	0,0059	9	0,9828	0,048	0,003
79	0,9964	0,006	0,0014	21	0,9964	0,0134	0,0032	77	0,9951	0,0507	0,0768
69	0,9964	0,0063	0,0015	41	0,9964	0,0163	0,0039	34	0,5294	0,0651	0,0742
30	0,9964	0,0064	0,0015	49	0,9855	0,017	0,0024	51	0,9993	0,0738	0,429
56	0,954	0,0065	0,0037	60	0,9964	0,02	0,0048	52	0,9776	0,0743	0,25
47	0,9964	0,0074	0,0018	39	0,9964	0,0209	0,005				

Tabla 9: Estadísticos de sumarización para los puntos de la fila

En este caso, el máximo es de 0,0743. Como $0,0743 < 0,15$ no elimino ninguna fila.

En segundo lugar analizo los resultados COLUMNA (Tabla 10)

Estadísticas de sumarización para los puntos de columna			
	Calidad	Mass	Inercia
CONSULTA_MOVIM	0	0	0
CONSULTA_SALDOS	0	0	0
ANUL_DEVOLUCION	0	0	0,0001
RECARGA_TELEFONO	0,0095	0,0001	0,0029
AJUSTE_DEBE_SALDO	0,0104	0,0002	0,0087
REINT_CAJERO_AJENO_NACION	0,1513	0,0061	0,0484
DISP_EFECT_OFICINA	0,9999	0,0965	0,4816
REINT_CAJERO_PROPIO	0,9971	0,1284	0,3392
COMPRA	0,9999	0,7686	0,1191

Tabla 10: Estadísticos de sumarización para los puntos de la columna

El valor máximo de estos datos es 0,7686 (correspondiente a “Compra”) que es mayor a $5 \times 0,03 = 0,15$, por lo tanto incluiré como suplementarias todas las categorías con valor inferior a 0,03, por lo tanto CONSULTA_MOVIM, CONSULTA_SALDOS, ANUL_DEVOLUCION, RECARGA_TELEFONO, AJUSTE_DEBE_SALDO, REINT_CAJERO_AJENO_NACION.

Si volvemos a hacer el análisis con estas modalidades como suplementarias obtenemos los siguientes resultados:

Lo primero que debo ver es la tabla de inercias y de la chi-cuadrado (Tabla 11)

Inercia y descomposición chi-cuadrado					
Valor singular	Inercia principal	Chi-cuadrado	Porcentaje	Porcentaje acumulado	12 24 36 48 60
0.84222	0.70933	4245709	59.38	59.38	*****
0.69656	0.48520	2904167	40.62	100.00	*****
Total	1.19452	7149876	100.00		
Grados de libertad = 104					

Tabla 11: Inercia y descomposición chi-cuadrado

En la Tabla 11 veo que con los dos primeros ejes tengo el 100% de la inercia recogida. Por lo tanto emplearé 2 dimensiones.

Analizo los resultados FILA

Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor
1	0	0	1
3	0	0	1
4	0	0	1
5	0	0	1
9	0	0	1
13	0	0	1
14	0	0	1
16	0	0	1
19	0	0	1
21	0	0	1
23	0	0	1
24	0	0	1
25	0	0	1
27	0	0	1

Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor
30	0	0	1
32	0	0	1
33	0	0	1
36	0	0	1
37	0	0	1
38	0	0	1
39	0	0	1
40	0	0	1
41	0	0	1
42	0	0	1
43	0	0	1
44	0	0	1
46	0	0	1
47	0	0	1

Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor
48	1	0	1
49	0	0	1
50	0	0	1
51	1	1	1
53	0	0	1
55	0	0	1
57	0	0	1
58	0	0	1
60	0	0	1
62	0	0	1
66	0	0	1
68	0	0	1
69	0	0	1
70	0	0	1

Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor
72	0	0	1
74	0	0	1
75	0	0	1
77	1	0	1
78	0	0	1
79	0	0	1
18	0	0	2
34	2	2	2
52	2	2	2
56	0	0	2
61	0	0	2

Tabla 12: Índices de las coordenadas que mejora contribuyan a la inercia para los puntos de la fila

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila		
	Dim1	Dim2		Dim1	Dim2		Dim1	Dim2		Dim1	Dim2
38	0,9941	0,0059	19	0,7486	0,2514	44	0,7486	0,2514	74	0,7486	0,2514
49	0,9399	0,0601	21	0,7486	0,2514	46	0,7486	0,2514	75	0,7486	0,2514
77	0,8452	0,1548	23	0,7486	0,2514	47	0,7486	0,2514	79	0,7486	0,2514
51	0,8409	0,1591	24	0,7486	0,2514	48	0,7486	0,2514	72	0,6567	0,3433
27	0,7801	0,2199	25	0,7486	0,2514	50	0,7486	0,2514	9	0,6562	0,3438
70	0,7555	0,2445	30	0,7486	0,2514	53	0,7486	0,2514	34	0,2654	0,7346
78	0,7501	0,2499	32	0,7486	0,2514	55	0,7486	0,2514	52	0,0556	0,9444
1	0,7486	0,2514	33	0,7486	0,2514	57	0,7486	0,2514	61	0,0467	0,9533
3	0,7486	0,2514	37	0,7486	0,2514	58	0,7486	0,2514	18	0,0052	0,9948
4	0,7486	0,2514	39	0,7486	0,2514	60	0,7486	0,2514	56	0,0001	0,9999
5	0,7486	0,2514	40	0,7486	0,2514	62	0,7486	0,2514	36	0	0
13	0,7486	0,2514	41	0,7486	0,2514	66	0,7486	0,2514			
14	0,7486	0,2514	42	0,7486	0,2514	68	0,7486	0,2514			
16	0,7486	0,2514	43	0,7486	0,2514	69	0,7486	0,2514			

Tabla 13: Cosenos cuadrados para los puntos de la fila

Con los resultados de la Tabla 12 y Tabla 13 obtengo la siguiente explicación de los ejes:

- La dimensión 1 está más influenciada por los consejeros GIL GARCÍA, MARÍA ELENA (38) y MARTÍNEZ VILCHES, RAMÓN (49)
- La dimensión 2 está más influenciada por los consejeros MORAL SANTÍN, JOSÉ ANTONIO (52), PRADILLO MORENO DE LA SANTA, RAFAEL (61), CABALLERO DOMÍNGUEZ, JOSÉ (18) y NAVALMORAL SANCHEZ, DOMINGO (56)

En segundo lugar analizo los resultados COLUMNA

Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de columna				Cosenos cuadrados para puntos de columnas		
	Dim1	Dim2	Mejor		Dim1	Dim2
COMPRA	1	0	1	COMPRA	0,8132	0,1868
DISP_EFECT_OFICINA	1	0	1	REINT_CAJERO_PROPIO	0,0988	0,9012
REINT_CAJERO_PROPIO	0	2	2	DISP_EFECT_OFICINA	0,8943	0,1057

Tabla 14: Resultados para los puntos de la columna

Con estas dos tablas obtengo la siguiente explicación de los ejes:

- La dimensión 1 está más influenciada por “Compra” y “Disposición de efectivo en oficina”.
- La dimensión 2 está más influenciada por el gasto “Reintegro en cajero propio”.

Los resultados de este análisis se detallan en el apartado 4.2. del presente documento.

ANÁLISIS CORRESPONDENCIAS SIMPLES ENTRE NOMBRE DE CONSEJERO Y TRAMO HORARIO

Lo primero que debo ver es la tabla de inercias y de la chi-cuadrado

Inercia y descomposición chi-cuadrado					
Valor singular	Inercia principal	Chi-cuadrado	Porcentaje	Porcentaje acumulado	16 32 48 64 80
0.37217	0.13851	1605132	81.49	81.49	*****
0.17735	0.03145	364510	18.51	100.00	*****
Total	0.16996	1969642	100.00		
Grados de libertad = 160					

Tabla 15: Inercia y descomposición Chi-cuadrado

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

En esta tabla veo que con los dos primeros ejes tengo el 100% de la inercia recogida.

A continuación veo si debo eliminar alguna variable

En primer lugar analizo los resultados FILA

Estadísticos de sumarización para los puntos de la fila				Estadísticos de sumarización para los puntos de la fila				Estadísticos de sumarización para los puntos de la fila				Estadísticos de sumarización para los puntos de la fila			
	Calidad	Mass	Inercia		Calidad	Mass	Inercia		Calidad	Mass	Inercia		Calidad	Mass	Inercia
36	0	0	0,0003	5	1	0,0041	0,0021	54	1	0,0106	0,0037	76	1	0,02	0,0011
62	1	0,0007	0,0002	24	1	0,0041	0,0087	39	1	0,0109	0,0072	1	1	0,0212	0,0015
3	0	0,0008	0,0001	74	1	0,0041	0,0006	46	1	0,011	0,0104	80	1	0,0212	0,0075
18	1	0,0015	0,0026	14	1	0,0045	0,0066	45	1	0,0113	0,0189	72	1	0,0213	0,0035
57	1	0,0015	0,0014	42	1	0,005	0,0044	58	1	0,0114	0,0008	70	1	0,0219	0,0066
37	1	0,0016	0,0003	4	1	0,0053	0,015	35	1	0,0116	0,0109	11	1	0,0226	0,0043
66	1	0,0016	0,0019	16	1	0,0054	0,0001	64	1	0,0118	0,0029	48	1	0,0235	0,004
53	1	0,0017	0,0037	25	1	0,0059	0,0018	8	1	0,0119	0,0311	22	1	0,0239	0,0089
55	1	0,0017	0,0018	33	1	0,0059	0,0051	13	1	0,012	0,0048	27	1	0,0245	0,0074
68	1	0,0017	0,0046	38	1	0,0062	0,0057	40	1	0,0128	0,0007	9	1	0,0248	0,006
23	1	0,0019	0,0021	12	1	0,0067	0,0076	50	1	0,0147	0,0059	77	1	0,0263	0,2041
61	1	0,002	0,001	21	1	0,0067	0,0022	19	1	0,015	0,002	59	1	0,0306	0,0072
44	1	0,003	0,0154	75	1	0,0067	0,007	29	1	0,015	0,0008	34	1	0,0336	0,001
79	1	0,0031	0,001	78	1	0,0068	0,0038	20	1	0,0152	0,0026	15	1	0,0365	0,0018
82	1	0,0031	0,0066	10	1	0,0084	0,0103	32	1	0,016	0,03	6	1	0,037	0,0073
30	1	0,0033	0,0332	41	1	0,0085	0,007	65	1	0,0163	0,0033	51	1	0,0383	0,2929
31	1	0,0033	0,0065	63	1	0,0085	0,0006	28	1	0,0166	0,0052	52	1	0,0386	0,0145
69	1	0,0033	0,0002	49	1	0,0086	0,0164	67	1	0,0177	0,0012	73	1	0,0494	0,0227
56	1	0,0034	0,0037	2	1	0,0092	0,0092	7	1	0,0179	0				
47	1	0,0038	0,0142	17	1	0,0092	0,0001	71	1	0,018	0,0113				
43	1	0,004	0,0012	60	1	0,0104	0,0223	26	1	0,0198	0,0051				

Tabla 16: Estadísticos de sumarización para los puntos de la fila

En este caso, el máximo es de 0,0494. Como $0,0494 < 0,15$, en este caso no elimino ninguna fila.

En segundo lugar analizo los resultados COLUMNA

Estadísticas de sumarización para los puntos de columna			
	Calidad	Mass	Inercia
N	1	0,0874	0,2361
M	1	0,309	0,5546
T	1	0,6035	0,2093

Tabla 17: Estadísticos de sumarización para los puntos de la fila

No elimino ninguna categoría ya que ningún valor de la masa se encuentra por debajo de 0,03.

Analizo los resultados FILA

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de fila				Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de fila				Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de fila				Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor		Dim1	Dim2	Mejor		Dim1	Dim2	Mejor		Dim1	Dim2	Mejor
2	1	0	1	59	0	0	1	11	0	2	2	40	0	0	2
3	0	0	1	62	0	0	1	12	0	2	2	42	0	2	2
5	0	0	1	63	0	0	1	13	0	0	2	43	0	0	2
9	0	0	1	65	0	0	1	14	0	2	2	44	0	2	2
10	1	0	1	66	0	0	1	16	0	0	2	46	0	2	2
15	0	0	1	68	0	0	1	20	0	2	2	47	0	2	2
17	0	0	1	70	0	0	1	21	0	0	2	48	0	0	2
18	0	0	1	71	1	0	1	23	0	0	2	49	0	2	2
19	0	0	1	73	1	1	1	25	0	0	2	50	0	0	2
22	1	0	1	74	0	0	1	26	0	2	2	52	2	2	2
24	1	0	1	75	0	0	1	27	0	2	2	54	0	0	2
29	0	0	1	76	0	0	1	28	0	2	2	56	0	0	2
32	1	1	1	77	1	0	1	30	2	2	2	58	0	0	2
38	0	0	1	79	0	0	1	31	0	0	2	60	2	2	2
41	0	0	1	82	0	0	1	33	0	2	2	61	0	0	2
45	1	0	1	1	0	0	2	34	0	0	2	64	0	0	2
51	1	0	1	4	0	2	2	35	0	2	2	67	0	0	2
53	0	0	1	6	0	2	2	36	0	0	2	69	0	0	2
55	0	0	1	7	0	0	2	37	0	0	2	72	0	0	2
57	0	0	1	8	2	2	2	39	0	0	2	78	0	2	2
												80	0	2	2

Tabla 18: Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de la fila

Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila			Cosenos cuadrados para los puntos de la fila		
	Dim1	Dim2		Dim1	Dim2		Dim1	Dim2		Dim1	Dim2
14	0	1	48	0,4502	0,5498	72	0,7261	0,2739	24	0,9566	0,0434
6	0,0001	0,9999	56	0,4505	0,5495	61	0,7376	0,2624	29	0,9574	0,0426
25	0,0161	0,9839	33	0,4891	0,5109	13	0,7798	0,2202	2	0,9584	0,0416
40	0,0387	0,9613	26	0,5064	0,4936	54	0,7997	0,2003	57	0,962	0,038
37	0,0552	0,9448	7	0,5073	0,4927	31	0,8025	0,1975	63	0,9656	0,0344
11	0,0706	0,9294	44	0,5106	0,4894	73	0,8318	0,1682	82	0,9779	0,0221
42	0,0732	0,9268	4	0,5127	0,4873	41	0,832	0,168	75	0,9791	0,0209
78	0,0966	0,9034	64	0,5318	0,4682	55	0,8367	0,1633	5	0,9849	0,0151
20	0,1006	0,8994	80	0,5367	0,4633	32	0,8594	0,1406	79	0,9849	0,0151
58	0,1673	0,8327	28	0,539	0,461	38	0,8618	0,1382	66	0,9892	0,0108
16	0,182	0,818	60	0,5451	0,4549	70	0,8683	0,1317	76	0,9911	0,0089
27	0,2022	0,7978	43	0,5547	0,4453	53	0,877	0,123	62	0,9936	0,0064
69	0,2041	0,7959	8	0,5782	0,4218	17	0,8837	0,1163	77	0,9937	0,0063
1	0,2236	0,7764	12	0,5954	0,4046	45	0,8967	0,1033	51	0,9938	0,0062
30	0,31	0,69	50	0,614	0,386	65	0,9021	0,0979	18	0,9957	0,0043
23	0,3284	0,6716	21	0,6473	0,3527	71	0,9113	0,0887	10	0,9963	0,0037
49	0,3292	0,6708	46	0,6568	0,3432	15	0,919	0,081	19	0,9984	0,0016
34	0,3571	0,6429	39	0,7084	0,2916	59	0,9479	0,0521	68	0,9999	0,0001
47	0,4306	0,5694	67	0,7223	0,2777	9	0,9507	0,0493	74	1	0
35	0,4489	0,5511	52	0,7253	0,2747	22	0,9523	0,0477	3	0	0
									36	0	0

Tabla 19: Cosenos cuadrados para los puntos de la fila

Con la Tabla 18 y Tabla 19 obtengo la siguiente explicación de los ejes:

- La dimensión 1 está más influenciada por los consejeros BLASCO BOSQUED, LUIS (14) y AMAT ROCA, MATÍAS (6)
- La dimensión 2 está más influenciada por los consejeros RIZALDOS GONZÁLEZ, ÁNGEL (68) y SÁNCHEZ CARLOS, SANTIAGO JAVIER (74)

En segundo lugar analizo los resultados COLUMNA

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

Indices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de columna				Cosenos cuadrados para puntos de columnas		
	Dim1	Dim2	Mejor		Dim1	Dim2
M	1	0	1	M	0,9955	0,0045
T	1	0	1	N	0,3684	0,6316
N	0	2	2	T	0,8403	0,1597

Tabla 20: Resultados para los puntos de la columna

Con la Tabla 20 obtengo la siguiente explicación de los ejes:

- La dimensión 1 está más influenciada por el gasto en el tramo horario de mañana
- La dimensión 2 está más influenciada por el tramo horario de noche

Los resultados de este análisis se detallan en el apartado 4.2. del presente documento.

ANALISIS CORRESPONDENCIAS SIMPLES ENTRE NOMBRE DE CONSEJERO Y DÍA DE LA SEMANA

Lo primero que debo ver es la tabla de inercias y de la chi-cuadrado

Inercia y descomposición chi-cuadrado					
Valor singular	Inercia principal	Chi-cuadrado	Porcentaje	Porcentaje acumulado	9 18 27 36 45
0.22142	0.04902	564467	46.70	46.70	*****
0.13261	0.01759	202479	16.75	63.45	*****
0.12008	0.01442	166015	13.73	77.18	*****
0.09914	0.00983	113169	9.36	86.54	*****
0.08575	0.00735	84671	7.00	93.55	****
0.08231	0.00677	78004	6.45	100.00	****
Total	0.10499	1208806	100.00		
Grados de libertad = 468					

Tabla 21: Inercia y descomposición chi-cuadrado

En la Tabla 21 veo que con los dos primeros ejes tengo el 63,45% de la inercia recogida.

A continuación veo si debo eliminar alguna variable

En primer lugar analizo los resultados FILA

Estadísticos de sumarización para los puntos de la fila				Estadísticos de sumarización para los puntos de la fila				Estadísticos de sumarización para los puntos de la fila				Estadísticos de sumarización para los puntos de la fila			
	Calidad	Mass	Inercia		Calidad	Mass	Inercia		Calidad	Mass	Inercia		Calidad	Mass	Inercia
36	0	0	0,0009	43	0,1444	0,0041	0,0064	54	0,8233	0,0106	0,004	26	0,2559	0,02	0,0081
62	0,1765	0,0007	0,0012	74	0,5766	0,0041	0,003	39	0,7518	0,0109	0,007	76	0,1414	0,0202	0,0092
3	0,4863	0,0008	0,0042	14	0,7179	0,0045	0,0033	46	0,6354	0,0111	0,0068	1	0,7298	0,0213	0,0176
18	0,3119	0,0015	0,0013	42	0,8354	0,005	0,012	45	0,7697	0,0114	0,0172	72	0,3772	0,0214	0,0095
57	0,0812	0,0015	0,0049	4	0,6223	0,0053	0,007	58	0,7001	0,0114	0,0034	80	0,1961	0,0214	0,025
37	0,5467	0,0016	0,0064	16	0,3734	0,0055	0,0018	35	0,6814	0,0117	0,0121	70	0,6507	0,0221	0,0122
66	0,2454	0,0016	0,0065	25	0,5696	0,0059	0,0053	64	0,6075	0,0119	0,0024	11	0,3221	0,0227	0,0029
53	0,4354	0,0017	0,0022	33	0,7011	0,0059	0,0134	8	0,9421	0,012	0,0245	48	0,4445	0,0236	0,0072
55	0,3145	0,0017	0,0023	38	0,1484	0,0062	0,0193	13	0,8503	0,0121	0,0053	22	0,2156	0,0241	0,0301
68	0,2516	0,0018	0,002	12	0,6836	0,0067	0,0162	40	0,3813	0,0129	0,0048	27	0,3876	0,0247	0,0026
23	0,2558	0,0019	0,0106	21	0,4183	0,0068	0,0036	50	0,8999	0,0148	0,0165	9	0,2809	0,0249	0,01
61	0,3174	0,002	0,0018	75	0,637	0,0068	0,0053	19	0,7226	0,0151	0,0135	77	0,9183	0,0265	0,0626
44	0,785	0,003	0,0023	78	0,8755	0,0069	0,0111	29	0,8369	0,0151	0,001	59	0,8452	0,0308	0,0622
79	0,2997	0,0031	0,0026	10	0,4911	0,0085	0,0017	20	0,0154	0,0153	0,0024	34	0,7479	0,0338	0,0126
69	0,2184	0,0033	0,0026	41	0,7926	0,0085	0,0111	32	0,7944	0,0161	0,0163	15	0,3359	0,0367	0,0402
30	0,2707	0,0034	0,0259	63	0,8148	0,0086	0,0208	65	0,2252	0,0164	0,0046	6	0,1696	0,0373	0,0348
56	0,3335	0,0034	0,0142	49	0,5957	0,0087	0,0124	28	0,8869	0,0167	0,0156	51	0,9343	0,0386	0,1149
47	0,7306	0,0038	0,0423	2	0,0452	0,0093	0,0024	67	0,5547	0,0178	0,0026	52	0,5852	0,0389	0,0017
5	0,5321	0,0041	0,0071	17	0,5927	0,0093	0,0034	7	0,8413	0,018	0,0051	73	0,824	0,0497	0,0308
24	0,1046	0,0041	0,0035	60	0,8831	0,0105	0,0183	71	0,6082	0,0181	0,0182				

Tabla 22: Estadísticos de sumarización para los puntos de la columna

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

En este caso, el máximo es de 0,0497. Como $0,0497 < 0,15$ no elimino ninguna fila.

En segundo lugar analizo los resultados COLUMNA

Estadísticas de sumarización para los puntos de columna			
	Calidad	Mass	Inercia
DOMINGO	0,8633	0,0811	0,2205
SABADO	0,7344	0,1401	0,1985
JUEVES	0,2949	0,1497	0,092
MIERCOLES	0,2609	0,1501	0,1119
LUNES	0,652	0,1503	0,152
MARTES	0,59	0,1609	0,1149
VIERNES	0,6818	0,1678	0,1102

Tabla 23: Resultados para los puntos de la columna

En este caso no elimino ninguna categoría ya que ningún valor de la masa se encuentra por debajo de 0,03.

Analizo los resultados FILA

Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor
2	0	0	1
4	0	0	1
5	0	0	1
6	0	0	1
8	1	1	1
10	0	0	1
12	1	0	1
13	0	0	1
14	0	0	1
16	0	0	1
17	0	0	1
19	1	0	1
21	0	0	1
27	0	0	1
28	1	1	1
29	0	0	1
33	1	1	1
35	1	0	1
37	0	0	1
39	0	0	1

Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor
41	1	0	1
42	1	0	1
45	1	0	1
47	1	0	1
49	0	0	1
50	1	1	1
51	1	0	1
52	0	0	1
53	0	0	1
59	1	1	1
60	1	0	1
63	1	1	1
64	0	0	1
67	0	0	1
68	0	0	1
71	1	1	1
73	1	1	1
75	0	0	1
78	1	0	1
80	0	0	1

Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor
1	2	2	2
3	0	2	2
7	0	0	2
9	0	0	2
11	0	0	2
15	2	2	2
18	0	0	2
20	0	0	2
22	0	2	2
23	0	2	2
24	0	0	2
25	0	0	2
26	0	0	2
30	0	2	2
32	0	2	2
34	0	2	2
36	0	0	2
38	0	0	2
40	0	0	2
43	0	0	2

Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de fila			
	Dim1	Dim2	Mejor
44	0	0	2
46	0	0	2
48	0	2	2
54	0	0	2
55	0	0	2
56	0	2	2
57	0	0	2
58	0	2	2
61	0	0	2
62	0	0	2
65	0	0	2
66	0	0	2
69	0	0	2
70	0	2	2
72	0	0	2
74	0	0	2
76	0	0	2
77	0	2	2
79	0	0	2

Tabla 24: Índices de las coordenadas que mejor contribuyan a la inercia para los puntos de la fila

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

Cosenos cuadrados para puntos de la fila			Cosenos cuadrados para puntos de la fila			Cosenos cuadrados para puntos de la fila			Cosenos cuadrados para puntos de la fila		
	Dim1	Dim2		Dim1	Dim2		Dim1	Dim2		Dim1	Dim2
77	0,0046	0,9137	50	0,6816	0,2183	22	0,1146	0,1009	39	0,7233	0,0285
58	0,0165	0,6836	30	0,0569	0,2138	76	0,0464	0,095	16	0,3454	0,028
44	0,131	0,6539	55	0,1221	0,1925	41	0,7039	0,0886	67	0,5276	0,0271
70	0,0098	0,6409	18	0,1269	0,185	8	0,8695	0,0726	68	0,2276	0,024
34	0,1632	0,5847	33	0,5208	0,1804	13	0,779	0,0713	6	0,1534	0,0162
74	0,0638	0,5127	65	0,047	0,1782	38	0,0805	0,0679	37	0,5312	0,0156
3	0,0453	0,441	66	0,071	0,1744	5	0,4709	0,0612	4	0,6099	0,0124
54	0,4045	0,4189	64	0,4599	0,1475	24	0,0471	0,0575	60	0,8731	0,01
32	0,3858	0,4086	63	0,668	0,1468	35	0,6261	0,0554	51	0,9254	0,0088
25	0,2609	0,3087	73	0,6776	0,1464	62	0,1261	0,0504	53	0,428	0,0074
56	0,06	0,2734	79	0,1566	0,1431	17	0,5437	0,0489	75	0,6297	0,0073
7	0,5692	0,272	43	0,0019	0,1425	19	0,6786	0,044	20	0,0096	0,0058
61	0,0495	0,2679	71	0,4667	0,1415	59	0,8027	0,0425	42	0,8304	0,005
48	0,1873	0,2572	72	0,2419	0,1354	45	0,7295	0,0401	10	0,4869	0,0043
1	0,4732	0,2566	15	0,202	0,1339	27	0,3494	0,0382	21	0,4149	0,0034
23	0,0017	0,2541	9	0,1527	0,1282	57	0,0438	0,0374	52	0,5822	0,003
46	0,392	0,2434	11	0,1989	0,1232	78	0,8411	0,0343	14	0,7158	0,0021
40	0,1464	0,235	69	0,0981	0,1203	12	0,6525	0,0311	2	0,045	0,0001
28	0,6551	0,2318	49	0,483	0,1128	47	0,6999	0,0307	36	0	0
29	0,6169	0,2201	26	0,1542	0,1017	80	0,1672	0,0289			

Tabla 25: Cosenos cuadrados para los puntos de la fila

Con la Tabla 24 y Tabla 25 obtengo la siguiente explicación de los ejes:

- La dimensión 1 está más influenciada por el consejero MORADO IGLESIAS, RICARDO (51)
- La dimensión 2 está más influenciada por el consejero TORRE MARTÍNEZ, ENRIQUE DE LA (77)

En segundo lugar analizo los resultados COLUMNA

Índices de las coordenadas que mejor contribuyan mejor a la inercia para los puntos de columna				Cosenos cuadrados para puntos de columnas		
	Dim1	Dim2	Mejor		Dim1	Dim2
DOMINGO	1	1	1	DOMINGO	0,7442	0,1191
LUNES	1	0	1	SABADO	0,7313	0,0031
SABADO	1	0	1	LUNES	0,6349	0,0171
JUEVES	0	2	2	MARTES	0,4209	0,169
MARTES	0	0	2	MIERCOLES	0,0665	0,1944
MIERCOLES	0	2	2	JUEVES	0,0513	0,2436
VIERNES	0	2	2	VIERNES	0,0063	0,6756

Tabla 26: Resultados para los puntos de la columna

Con la Tabla 26 obtengo la siguiente explicación de los ejes:

- La dimensión 1 está más influenciada por el gasto en domingos y sábados
- La dimensión 2 está más influenciada por el gasto en viernes

Los resultados de este análisis se detallan en el apartado 4.2. del presente documento.

3.2.4.2. Análisis clúster

Se han empleado tres técnicas de clustering diferentes:

- Análisis clúster jerárquico.
- Análisis cluster no jerárquico (k-means)
- Análisis clúster difuso (fuzzy)

ANÁLISIS CLUSTER JERÁRQUICO

En este tipo de análisis, se parte de tantos grupos como observaciones. Estas observaciones se van agrupando configurando una estructura arborescente llamada dendrograma, y se van formando agrupaciones hasta llegar a un único grupo que engloba a todos los individuos.

Se busca formar grupos cuyos individuos sean lo más parecido posible entre ellos y sean lo más diferentes posible del resto de los grupos.

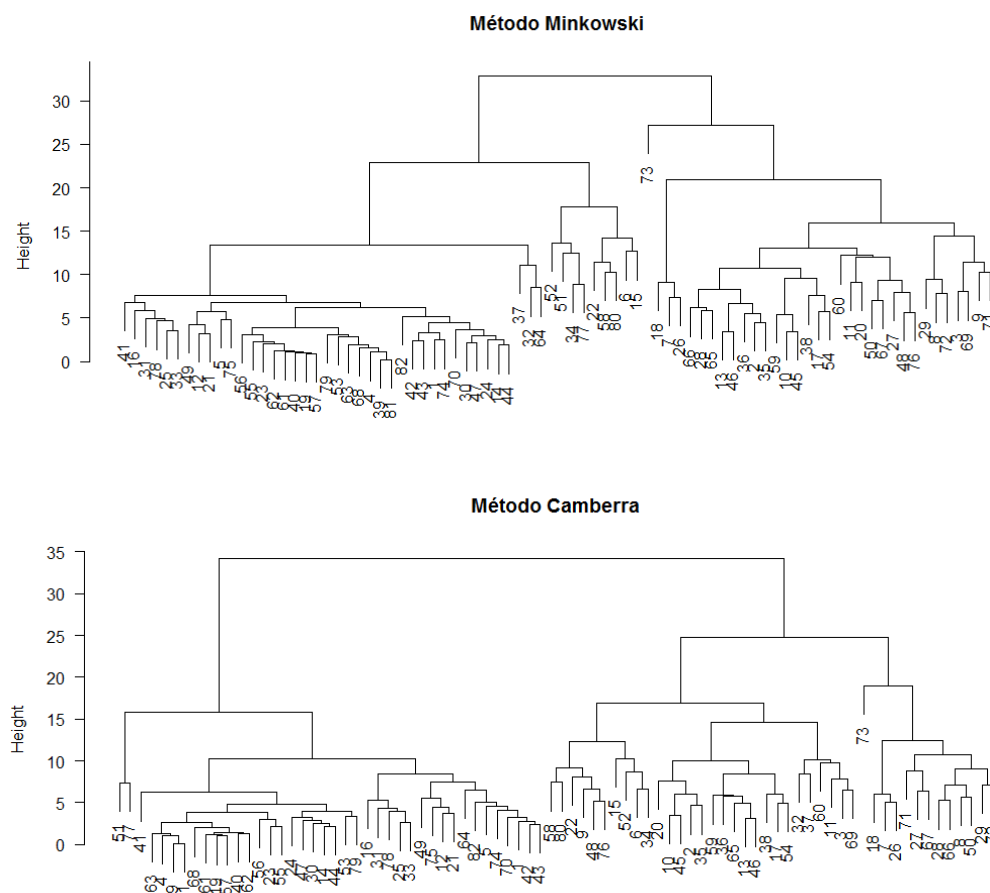
Las principales fases de este análisis son:

- Cálculo de la matriz de distancias: necesaria para medir la semejanza entre dos observaciones. En el presente estudio se ha valorado la utilización de tres distancias diferentes:

- Distancia de Minkowski: $\delta_{m_q}(i, j) = (\sum_{k=1}^p |x_{ik} - x_{jk}|^q)^{\frac{1}{q}}$ $q > 0$
- Distancia Euclídea: $\delta_E^2(i, j) = \sum_{k=1}^p (x_{ik} - x_{jk})^2$
- Distancia de Canberra: $\delta_C(i, j) = \sum_{k=1}^p \frac{|x_{ik} - x_{jk}|}{|x_{ik}| + |x_{jk}|}$

Siendo en todos los casos $x'_i = (x_{i1}, \dots, x_{ip})$ y $x'_j = (x_{j1}, \dots, x_{jp})$ las observaciones de dos objetos o individuos i, j , resultado de medir P variables X_1, X_2, \dots, X_p sobre ellos.

Si con esta matriz de distancias, empleamos el algoritmo del “enlace completo o del vecino más alejado” (que es el método por defecto empleado por el software R 3.1.3. ®), los dendrogramas obtenidos con cada una de las distancias se muestran en la Figura 1.



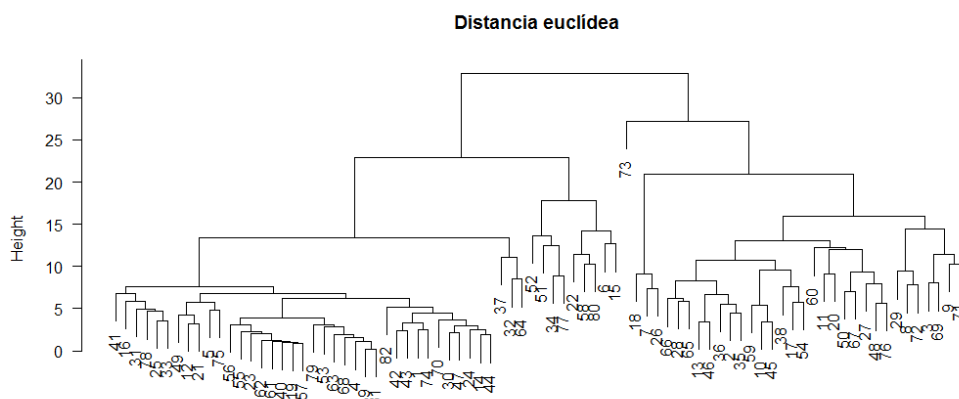
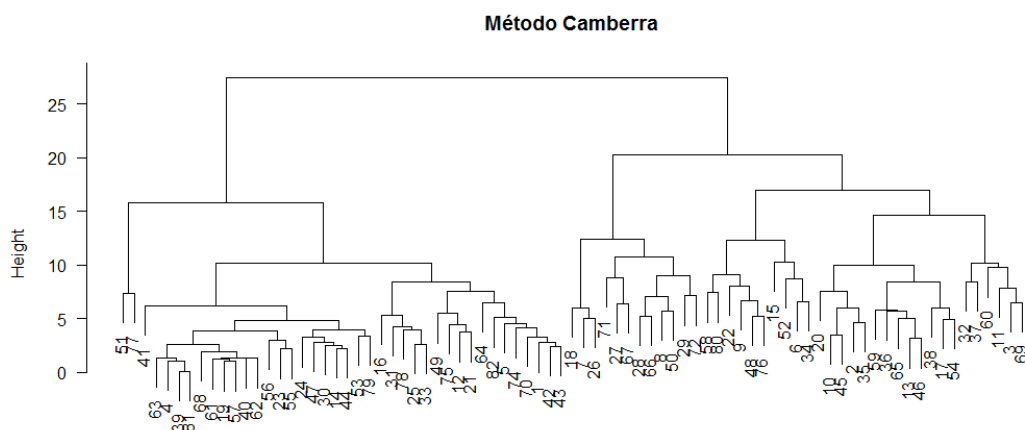
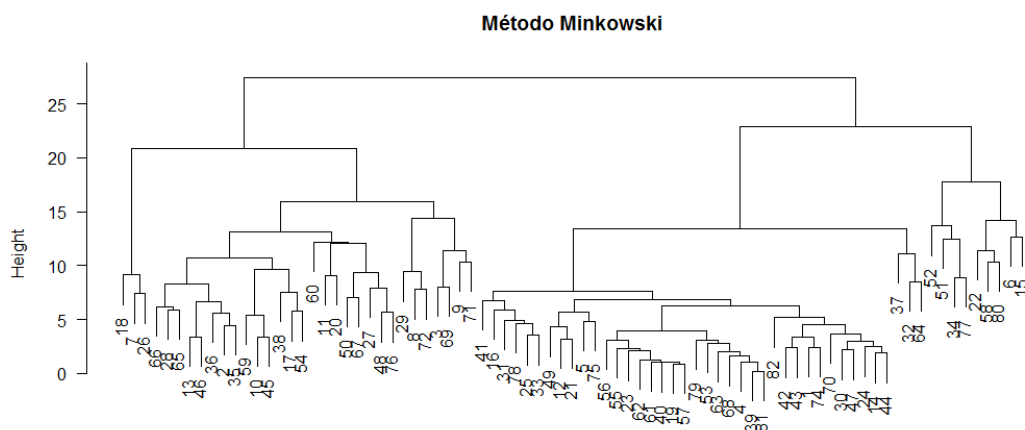


Figura 1. Dendrogramas clúster jerárquico

En el algoritmo del “enlace completo o vecino más alejado” la distancia elegida entre dos clúster es: $d(C_k, C_{k'}) = \max d(u_{k,i}, u_{k',i'})$ con $i=1,...,n_k$ $i'=1,...,n_{k'}$, teniendo una agrupación denominada C_k formada por las observaciones $u_{k,1},...,x_{k,n_k}$ y otra agrupación $C_{k'}$ formada por las observaciones $u_{k',1},...,x_{k',n_{k'}}$

En cualquiera de los tres casos vemos cómo la observación 73, correspondiente a Ildefonso José Sánchez Barcoj, queda fuera de cualquiera de las agrupaciones. Por esta razón, la vamos a eliminar del conjunto de datos y posteriormente realizaremos su estudio por separado.

Eliminando esta observación, los dendrogramas quedan según la Figura 2.



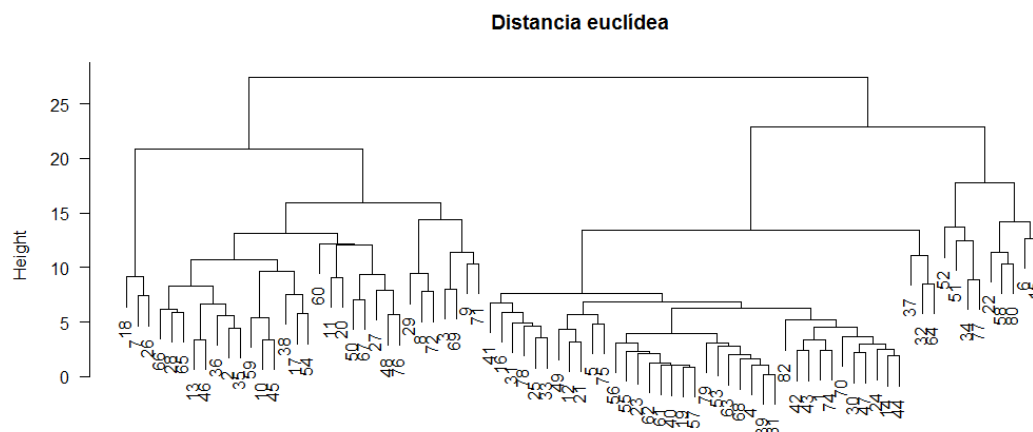


Figura 2. Dendrogramas clúster jerárquico

Tomamos la distancia euclídea para continuar el estudio.

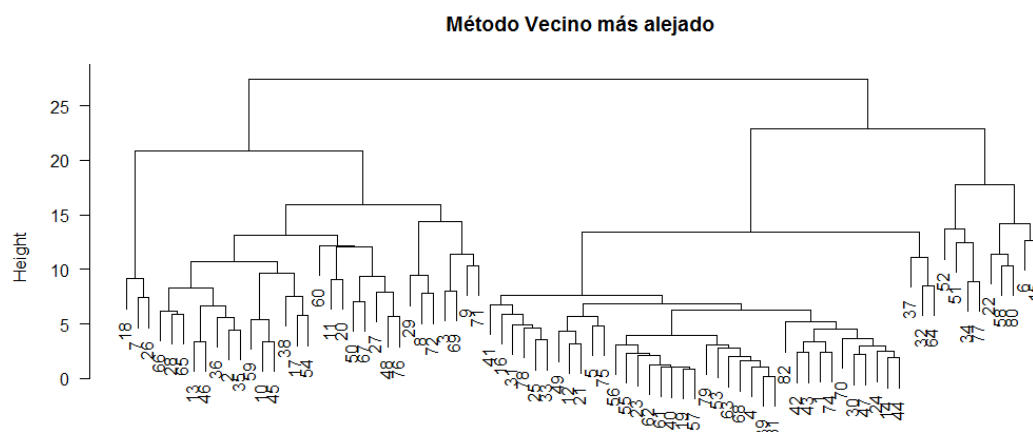
- Elección del algoritmo de clasificación jerárquica: una vez hemos decidido emplear la distancia euclídea, aplicamos diferentes métodos de clasificación para quedarnos con el más adecuado. Si suponemos que tenemos una agrupación denominada C_k formada por las observaciones $u_{k,1}, \dots, u_{k,n_k}$ y otra agrupación $C_{k'}$ formada por las observaciones $u_{k',1}, \dots, u_{k',n_{k'}}$

Los métodos que se han estudiado son los siguientes:

- Método del enlace completo o del vecino más alejado: la distancia elegida entre dos clúster es: $d(C_k, C_{k'}) = \max d(u_{k,i}, u_{k',i'})$ con $i=1, \dots, n_k$ $i'=1, \dots, n_{k'}$
- Método de Ward o de la mínima varianza: la distancia elegida es la suma de cuadrados de los grupos que se pueden formar en cada paso de la forma:

$$d(C_k, C_{k'}) = \frac{\sum_{j=1}^p (\bar{x}_{k,j} - \bar{x}_{k',j})^2}{\frac{1}{c_k} + \frac{1}{c_{k'}}}$$
- Método del vecino más cercano: la distancia entre dos grupos es la distancia entre las dos observaciones más cercanas, pertenecientes cada una de ellas a un grupo distinto. De este modo tenemos: $d(C_k, C_{k'}) = \min d(u_{k,i}, u_{k',i'})$ con $i=1, \dots, n_k$ $i'=1, \dots, n_{k'}$

Los dendrogramas de cada uno de los métodos se muestran en la Figura 3.



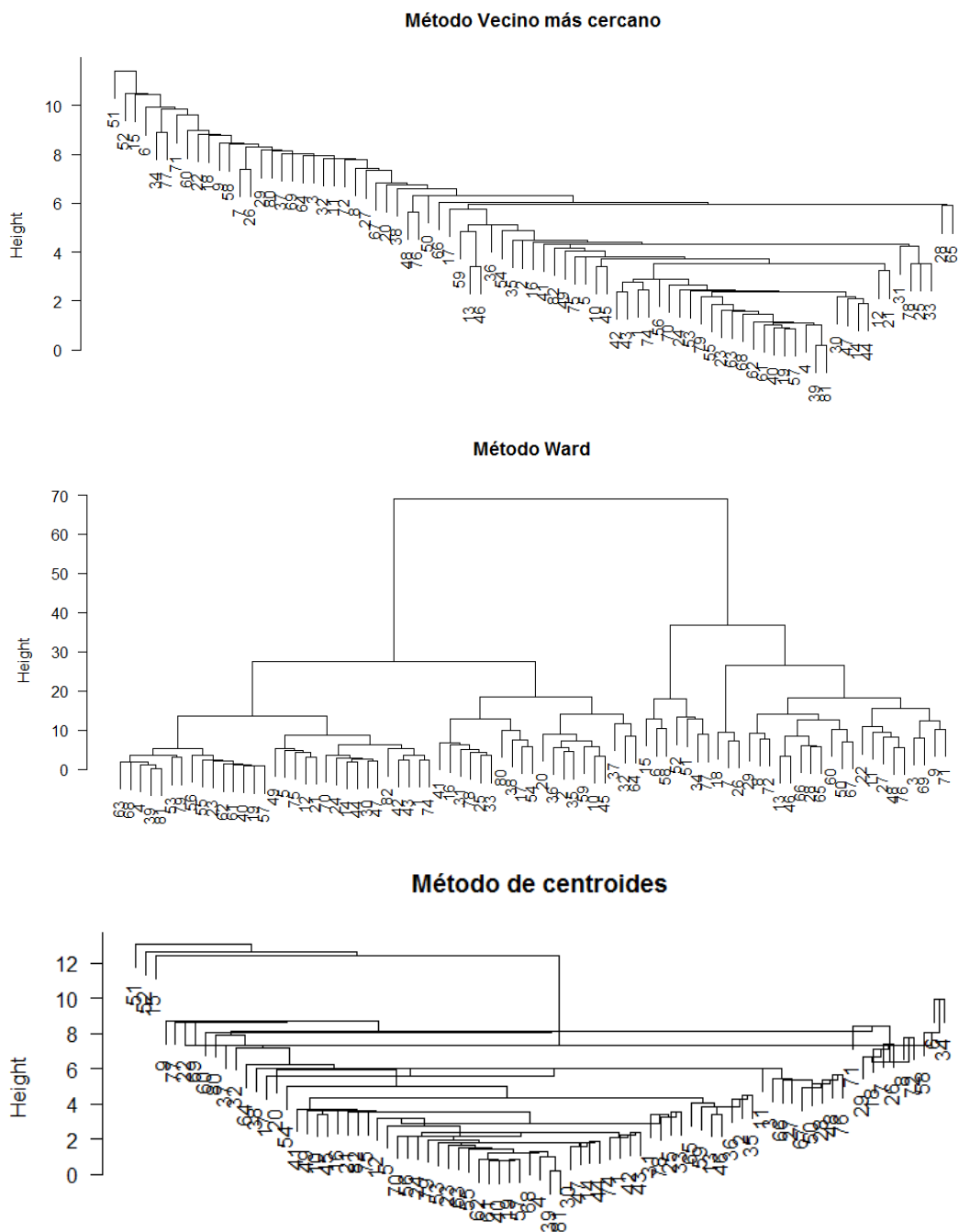


Figura 3. Dendrogramas clúster jerárquico III

A la vista de estos dendrogramas, podríamos quedarnos con el método del “vecino más alejado” o con el “método de Ward”. En este caso se ha decidido emplear el “método de Ward”

- Elección de número de clusters: tras una exploración gráfica del dendrograma, vemos que podríamos emplear 3, 4 o 5 clusters.

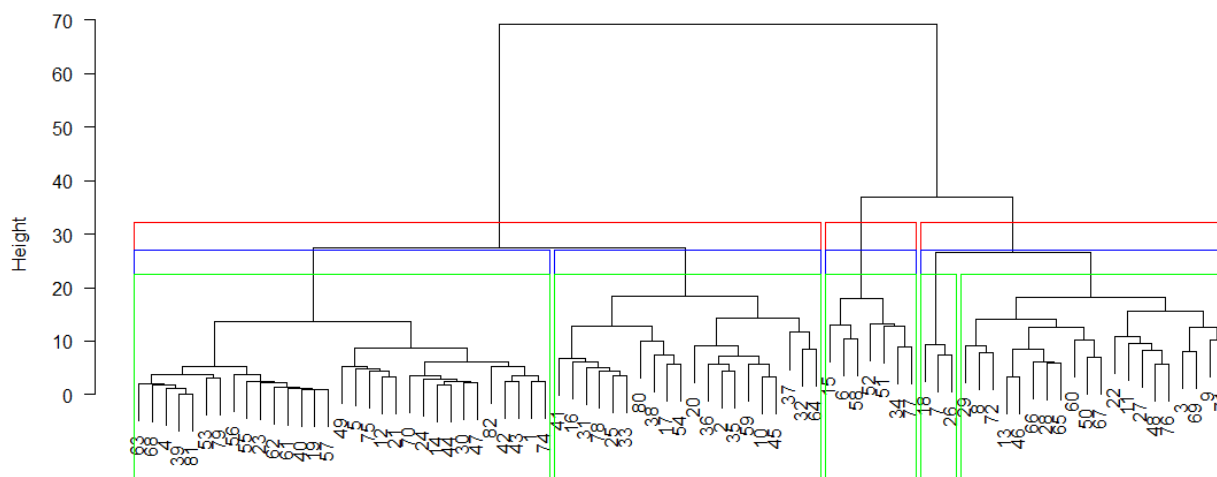


Figura 4. Dendrogramas clúster jerárquico

Decidimos quedarnos con **4 clusters**, ya que con 3 nos quedaría un clúster con demasiadas observaciones en comparación con los otros, y con 5 clusters uno de ellos nos quedaría demasiado pequeño.

ANÁLISIS CLUSTER NO JERÁRQUICO (K-MEANS)

En este tipo de algoritmo se parte de unas medias arbitrarias para cada clúster, mediante pruebas sucesivas contrasta el efecto que sobre la varianza residual tiene la asignación de cada una de las observaciones a cada grupo. El valor de mínima varianza determina la asignación, este paso se repite hasta que todas las observaciones hayan sido asignadas. A continuación se determina el centroide de todos los grupos y se repite el proceso de asignación nuevamente desde la primera a la última observación. Las iteraciones se suceden hasta que los centroides convergen.

En primer lugar se ha determinado el número de clusters óptimo. Para poder determinarlo se ha usado el algoritmo de K-Means de forma repetida (iterada), aumentando la cantidad de grupos (clúster) en cada ejecución y obteniendo el error intracluster en cada iteración, donde este error calcula la distancia entre cada observación y el clúster al cual pertenece.

Con esta técnica (conocida como SSE ó Sum of Squares Error) se obtiene la siguiente gráfica que permite identificar la cantidad ideal de clúster o grupos.

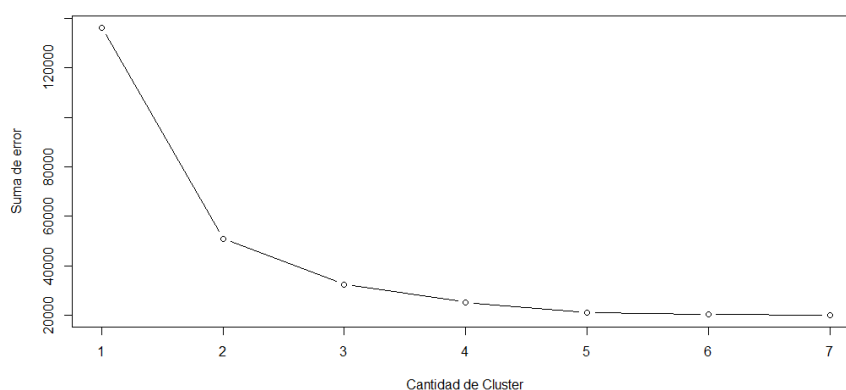


Figura 5. Sum of Squares Error

A la vista del gráfico, podríamos quedarnos con 4 o 5 clusters. Decidimos quedarnos con 4 clúster.

Por tanto, el modelo que vamos a emplear será **“Segmentación k-means en 4 clusters tomando la matriz de distancias euclídeas”**, exceptuando la observación correspondiente a Ildefonso José Sánchez Barcoj que estudiaremos por separado.

Los resultados de la aplicación de este modelo en los datos se detallan en el apartado 4.2. del presente documento.

ANÁLISIS CLUSTER DIFUSO (FUZZY CLUSTERING)

La necesidad de empleo de los algoritmos difusos surge por la necesidad de resolver una deficiencia del agrupamiento exclusivo, que considera que cada elemento se puede clasificar inequívocamente con los elementos de su clúster, y que por lo tanto no se asemeja al resto de elementos.

El fuzzy clustering es una clase de análisis clustering donde cada individuo tiene un grado de pertenencia difuso a los diferentes clúster, es decir, no pertenece a un grupo y deja de tener relación con los demás, sino que tiene un porcentaje de pertenencia a cada clúster.

Esto se logra representando la similitud entre un elemento y un grupo por una función (función de pertenencia) que toma valores entre 0 y 1. Los valores cercanos a 1 indican mayor similitud mientras que los cercanos a 0 indican menor similitud.

Si aplicamos el fuzzy clustering sobre los datos con un total de 4 clusters (en este caso lo hacemos con todas las observaciones, no eliminando la correspondiente a Ildefonso José Sánchez Barcoj), vemos que los clusters quedan formados por 16, 12, 28 y 26 observaciones respectivamente.

Viendo la Tabla 27 con los grados de pertenencia de las observaciones a cada uno de los clusters:

- La observación 1 pertenece en un 33% al clúster 1, en un 18% al clúster 2, en un 18% al clúster 3 y en un 32% al clúster 4, por lo que finalmente el algoritmo la categoriza como clúster 1.
- La observación 2 pertenece en un 28% al clúster 1, al clúster 2, y al clúster 3 y en un 16% al clúster 4, y finalmente el algoritmo la categoriza como clúster 2.
- La observación 3 pertenece en un 20% al clúster 1, en un 34% al clúster 2, en un 34% al clúster 3 y en un 12% al clúster 4, y finalmente el algoritmo la categoriza como clúster 3.
- ...y así con todas las observaciones.

	V1	V2	V3	V4		V1	V2	V3	V4		V1	V2	V3	V4		V1	V2	V3	V4
1	0,33	0,18	0,18	0,32	22	0,21	0,33	0,33	0,13	43	0,32	0,17	0,17	0,34	64	0,32	0,23	0,23	0,22
2	0,28	0,28	0,28	0,16	23	0,26	0,14	0,14	0,47	44	0,26	0,13	0,13	0,47	65	0,22	0,32	0,32	0,13
3	0,20	0,34	0,34	0,12	24	0,30	0,15	0,15	0,40	45	0,30	0,25	0,25	0,20	66	0,20	0,34	0,34	0,11
4	0,23	0,13	0,13	0,50	25	0,33	0,21	0,21	0,25	46	0,23	0,32	0,32	0,13	67	0,18	0,36	0,36	0,11
5	0,31	0,17	0,17	0,34	26	0,19	0,35	0,35	0,12	47	0,29	0,15	0,15	0,41	68	0,23	0,12	0,12	0,52
6	0,21	0,33	0,33	0,13	27	0,18	0,35	0,35	0,11	48	0,19	0,35	0,35	0,11	69	0,21	0,33	0,33	0,13
7	0,19	0,35	0,35	0,12	28	0,18	0,36	0,36	0,10	49	0,32	0,22	0,22	0,23	70	0,29	0,15	0,15	0,41
8	0,22	0,32	0,32	0,14	29	0,19	0,35	0,35	0,12	50	0,19	0,35	0,35	0,11	71	0,20	0,34	0,34	0,12
9	0,19	0,35	0,35	0,11	30	0,27	0,15	0,15	0,43	51	0,26	0,27	0,27	0,20	72	0,20	0,34	0,34	0,12
10	0,32	0,24	0,24	0,19	31	0,31	0,16	0,16	0,36	52	0,22	0,31	0,31	0,15	73	0,20	0,34	0,34	0,13
11	0,21	0,33	0,33	0,12	32	0,25	0,30	0,30	0,15	53	0,25	0,13	0,13	0,48	74	0,33	0,18	0,18	0,31
12	0,35	0,20	0,20	0,25	33	0,33	0,21	0,21	0,25	54	0,28	0,27	0,27	0,17	75	0,34	0,21	0,21	0,24
13	0,22	0,33	0,33	0,12	34	0,21	0,33	0,33	0,13	55	0,24	0,12	0,12	0,52	76	0,18	0,35	0,35	0,11
14	0,30	0,15	0,15	0,39	35	0,26	0,30	0,30	0,15	56	0,27	0,15	0,15	0,44	77	0,27	0,27	0,27	0,20
15	0,19	0,34	0,34	0,12	36	0,23	0,32	0,32	0,13	57	0,23	0,12	0,12	0,54	78	0,32	0,21	0,21	0,25
16	0,34	0,21	0,21	0,24	37	0,27	0,28	0,28	0,17	58	0,20	0,34	0,34	0,12	79	0,28	0,14	0,14	0,44
17	0,26	0,30	0,30	0,15	38	0,24	0,31	0,31	0,14	59	0,27	0,27	0,27	0,18	80	0,23	0,31	0,31	0,15
18	0,19	0,34	0,34	0,12	39	0,23	0,13	0,13	0,50	60	0,22	0,32	0,32	0,13	81	0,24	0,14	0,14	0,49
19	0,23	0,12	0,12	0,53	40	0,23	0,12	0,12	0,53	61	0,23	0,12	0,12	0,52	82	0,32	0,18	0,18	0,33
20	0,26	0,29	0,29	0,16	41	0,31	0,19	0,19	0,32	62	0,24	0,13	0,13	0,51					
21	0,34	0,21	0,21	0,24	42	0,34	0,18	0,18	0,31	63	0,24	0,13	0,13	0,50					

Tabla 27. Grados de pertenencia de las observaciones en el clúster difuso

Los resultados de la aplicación de este modelo en los datos se detallan en el apartado 4.2. del presente documento.

3.2.5. Assesment (Evaluar)

Las conclusiones y resultados obtenidos tanto en el análisis descriptivo como del análisis de correspondencias y de los 3 modelos de análisis clúster realizados se detallan en los apartados 4 y 5 del presente documento.

3.3. Software

Para la realización de la investigación se ha empleado el siguiente software estadístico:

- **SAS Enterprise Guide 6.1**®:
 - Para la recogida de los datos y su unión en un único archivo
 - Para la exploración numérica de los datos
 - Para la realización del análisis de correspondencias
- **R 3.1.3.**® y **RStudio**®:
 - Para la exploración gráfica de los datos
 - Para la realización del análisis del análisis clúster jerárquico, análisis clúster no jerárquico (k-means) y análisis clúster difuso.

Asimismo, se ha empleado Microsoft Word para la redacción del informe y Microsoft Power Point para la elaboración de la presentación final.

4. DESARROLLO DEL TRABAJO Y PRINCIPALES RESULTADOS

4.1. Cronograma

4.1.1. Fases o etapas de la investigación

Se detallan a continuación las distintas fases en que se divide el presente Trabajo Fin de Máster:

1. Depuración de los datos: se ha obtenido de la web un fichero por cada uno de los 82 consejeros y a partir de ellos se ha generado un único fichero con todos los datos. Se han corregido errores de digitalización, datos incoherentes, valores faltantes,... que podrían deteriorar la calidad de la información y en consecuencia, la calidad de las decisiones que se tomen.
2. Análisis estadístico de los datos:
 - a) Análisis descriptivo de los datos: se ha realizado una descripción tanto numérica como gráfica de los datos, tanto de forma global como individualizada para cada uno de los consejeros.
 - b) Estudio analítico de los datos: se ha realizado tanto un análisis clúster como un análisis clúster difuso, realizando previamente en ambos casos una estandarización de los datos y un análisis de componentes principales para determinar las variables que más influyen en la segmentación....
3. Interpretación de los datos.
4. Conclusiones y elaboración del informe final.
5. Presentación y exposición del trabajo final.

4.1.2. Periodo, lugar y materiales para el desarrollo de la investigación

El presente Trabajo Fin de Máster se presenta en la convocatoria de **Noviembre 2015**, contando por tanto con una duración de **11 meses**.

Para el desarrollo de la investigación se ha dispuesto de los siguientes emplazamientos:

- Sala de Informática de la Facultad de Estudios Estadísticos de la Universidad Complutense de Madrid
- Biblioteca de la Facultad de Estudios Estadísticos de la Universidad Complutense
- Sala de estudio en mi casa

El material con el que se contará para el desarrollo del trabajo será, a grandes rasgos, el que se detalla a continuación:

- Material didáctico de las diferentes asignaturas del Máster en Minería de Datos e Inteligencia de Negocio
- Recursos disponibles en la biblioteca de la Facultad

- Recursos disponibles en la red: revistas, publicaciones, bases de datos documentales, buscadores científicos...
- Ordenadores disponibles en la sala de informática de la Facultad así como ordenador portátil propio.

Se puede consultar más detalladamente la bibliografía empleada en la investigación en el apartado 6 del presente documento.

4.1.3. Cronograma

Se incluye a continuación la distribución en el tiempo de las distintas fases llevadas a cabo en el presente Trabajo Fin de Máster.

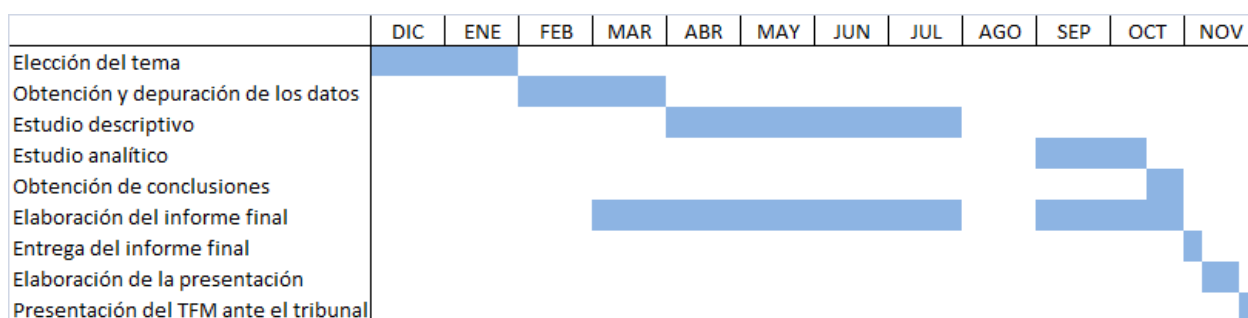


Figura 5. Cronograma

4.2. Principales resultados

4.2.1. Exploración de los datos. Resultados destacables del análisis descriptivo

Del análisis descriptivo realizado sobre los datos de gasto totales (11.588.629€) realizados por los 82 consejeros, destacamos como información más relevante:

CONSIDERANDO LA VARIABLE TIPO DE GASTO

	TIPO DE GASTO			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
RESTAURANTES RESTO	18.205	24,11%	2.165.402	18,69%
CAJEROS AUTOMATICOS	2.544	3,37%	1.230.712	10,62%
EL CORTE INGLES	4.874	6,46%	975.722	8,42%
AGENCIAS BANCARIAS(ANTICIPO VENTANILLA)	657	0,87%	835.561	7,21%
HOTELES 4 Y 5 ESTRELLAS,BALNEARIOS,CAMPI	1.678	2,22%	449.423	3,88%
HIPERCOR SUPERMERCADOS EL CORTE INGLES	2.209	2,93%	423.220	3,65%

Tabla 28. Análisis de la variable Tipo de Gasto

El mayor gasto se realizó en “Restaurantes Resto” suponiendo un 18,69% de la cantidad total gastada. El siguiente gasto más importante fue la extracción de efectivo en “Cajeros automáticos” (un 10,62%), seguido por compras en “El Corte Inglés” (8,42%), Anticipos en Ventanilla (7,21%)...

Si desglosamos algunos de estos “tipos de gasto” por el nombre del comercio donde se realizaron, podemos ver:

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

RESTAURANTES RESTO: en este caso vemos que el gasto está muy atomizado entre los diferentes restaurantes, no siendo mayor al 2% en ninguno de los casos. Por otra parte, el 6,35% de la cantidad gastada en este concepto, se realizó en restaurantes de los que desconocemos su nombre.

AGENCIAS DE VIAJES: las agencias en las que mayor gasto se hizo se muestran en la siguiente tabla:

	AGENCIAS DE VIAJES			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
MILLAN TRAVEL SA	47	10,63%	48.324	12,42%
VIAJES TIKAL S.A.	35	7,92%	33.367	8,57%
VIAJES CATAI SA	3	0,68%	25.696	6,60%

Tabla 29. Análisis de Agencias de Viajes

El gasto mayor se realizó en la agencia de viajes “Millan Travel S.A.” con un 12,42% de la cantidad total gastada, seguida de “Viajes Tikal”, “Viajes Catai”...

CONFECCIÓN TEXTIL EN GENERAL

	CONFECCION TEXTIL EN GENERAL			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
YUSTY	48	3,41%	31.882	8,68%
	83	5,89%	17.450	4,75%

Tabla 30. Análisis de Confección Textil en general

La tienda en que mayor gasto se realizó (8,68%) fue en Yusty. El 4,75% de la cantidad gastada en este concepto se realizó en tiendas de las que no disponemos su nombre.

GASOLINERAS: las gasolineras en las que mayor gasto se hizo se muestran en la siguiente tabla:

	GASOLINERAS			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
E.S. EXPLOTACIONES IND	241	4,16%	12.576	4,02%
EST SERV SHELL	153	2,64%	6.756	2,16%
E S MONTECLARO S A	97	1,67%	5.853	1,87%

Tabla 31. Análisis de Gasolineras

El gasto en gasolineras se encuentra bastante atomizado, destacando únicamente por encima de las demás la “E.S. Explotaciones Industriales” con 4,02% del gasto total en gasolineras.

CONSIDERANDO LA VARIABLE OPERACIÓN

Si mostramos la tabla de frecuencias de esta variable y el gasto realizado en cada tipo de operación obtenemos lo siguiente:

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

	OPERACIÓN			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
COMPRA	71.654	94,97%	9.712.133,55	83,81%
REINTEGRO EN CAJERO PROPIO	2.346	3,11%	1.185.780,00	10,23%
DISPOSICION EFECTIVO OFICINA	598	0,79%	835.560,80	7,21%
REINTEGRO EN CAJERO AJENO NACIONAL	134	0,18%	42.725,51	0,37%
RECARGA TELEFONO GSM - CARGA EN F	59	0,08%	2.805,00	0,02%

Tabla 32. Análisis de la variable Operación

Podemos ver cómo las operaciones más importantes tanto por frecuencia como por cantidad gastada son **“Compra”**, **“Reintegro en Cajero Propio”**, **“Disposición efectivo en oficina”** y en menor medida **“Reintegro en cajero Ajeno Nacional”**

Si realizamos un estudio en profundidad de estas operaciones.

COMPRA

Para las operaciones de **“Compra”** calculamos los siguientes estadísticos:

- La cantidad máxima gastada es de 16.921,7€
- La cantidad media gastada es de 135,54€

Si tramificamos la cantidad gastada en los siguientes intervalos:

- Compra menor a 150€
- Compra entre 150 y 500€
- Compra mayor a 500€

Y estudiamos dónde se han efectuado esas **“Compras”** obtenemos los siguientes datos:

	<150€				150-500€				>500€				TOTAL			
	Cantidad (€)	%	Cantidad	Frec. % Frec.	Cantidad (€)	%	Cantidad	Frec. % Frec.	Cantidad (€)	%	Cantidad	Frec. % Frec.	Cantidad (€)	%	Cantidad	Frec. % Frec.
RESTAURANTES RESTO	1.014.663	47%	13.570	75%	994.553	46%	4.409	24%	157.243	7%	212	1%	2.166.458	22%	18.191	25%
EL CORTE INGLES	195.629	20%	3.415	72%	271.127	27%	1.039	22%	528.655	53%	317	7%	995.411	10%	4.771	7%
HOTELES 4 Y 5 ESTRELLAS, BALNEARIOS, CAMPI	65.987	15%	993	60%	125.096	28%	478	29%	261.247	58%	190	11%	452.329	5%	1.661	2%
HIPERCOR SUPERMERCADOS	60.571	14%	1.190	54%	202.087	47%	808	37%	164.561	39%	195	9%	427.220	4%	2.193	3%
EL CORTE INGLES	133.129	32%	1.436	57%	242.866	59%	1.047	41%	39.012	9%	54	2%	415.006	4%	2.537	4%
RESTAURANTES DE 4 Y 5 TENEDORES	8.123	2%	140	32%	30.210	7%	100	23%	370.004	91%	192	44%	408.338	4%	432	1%
AGENCIAS DE VIAJES	58.364	16%	737	53%	136.233	37%	492	35%	173.400	47%	174	12%	367.998	4%	1.403	2%
CONFECCION TEXTIL EN GENERAL	299.865	96%	5.752	99%	7.069	2%	24	0%	6.310	2%	9	0%	313.245	3%	5.785	8%
GASOLINERAS	1.336.989	32%	28.541	82%	1.217.379	29%	4.794	14%	1.611.762	39%	1.346	4%	4.166.129	43%	34.681	48%
RESTO	3.173.320	33%	55.774	78%	3.226.620	33%	13.191	18%	3.312.194	34%	2.689	4%	9.712.134	100%	71.654	100%
Total general																

Tabla 33. Análisis de la variable Compra

Como información más destacable de la Tabla 33 tenemos:

- En concepto de compra se gastaron 9.712.134€, lo que supone el 83,81% de la cantidad gastada con las tarjetas Black, siendo como se ha indicado anteriormente el gasto más importante.
- Las anotaciones realizadas como **“Restaurantes Resto”** suponen el 22% de la cantidad total (con una frecuencia de utilización del 25%). Es el gasto mayor tanto en cantidad como en frecuencia.

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

- El segundo gasto más importante en cuanto a cantidad es el realizado en el “**El Corte Inglés**”, con un 10 % del gasto. Dentro de este 10%, el 53% del gasto corresponde a operaciones de más de 500€.
- El cuarto gasto más importante en cuanto a cantidad (4%) se realizó en concepto de “**Hipermercados El Corte Inglés**”, suponiendo un total de 427.220€. De este gasto, el 47% se realizó en operaciones de entre 150-500€ y el 39% en operaciones de más de 500€.
- El segundo gasto más importante en cuanto a frecuencia (8%) es el realizado en “**Gasolineras**”, suponiendo el 3% de la cantidad total. De este gasto, el 2% (6.310€) se realizó en 9 operaciones de más de 500€.
- El gasto en “**Confección textil en general**” supone el 4% del gasto total. De este gasto, el 47% se realizó en operaciones de más de 500€, suponiendo un total de 173.400€.

REINTEGRO EN CAJERO PROPIO

Para las operaciones de “Reintegro en cajero propio” calculamos los siguientes estadísticos:

- La cantidad máxima sacada es de 1.000€
- La cantidad media sacada es de 505,45€
- La cantidad mínima sacada es de 20€

Si tramificamos la cantidad gastada en intervalos de 100€:

	Cantidad (€)	% cantidad	Frecuencia	% frecuencia
<100 €	780	0,07%	16	0,68%
Entre 100 y 200 €	10.380	0,88%	88	3,75%
Entre 200 y 300 €	6.560	0,55%	32	1,36%
Entre 300 y 400 €	94.540	7,97%	314	13,38%
Entre 400 y 500 €	6.400	0,54%	16	0,68%
Entre 500 y 600 €	428.000	36,09%	856	36,49%
Entre 600 y 700 €	576.020	48,58%	960	40,92%
Entre 700 y 800 €	700	0,06%	1	0,04%
Entre 800 y 900 €	0	0,00%	0	0,00%
>900 €	62.400	5,26%	63	2,69%
Total general	1.185.780	100%	2346	100%

Tabla 34. Análisis de la operación Reintegro en Cajero Propio

Como información más destacable de la Tabla 34 tenemos:

- En concepto de “Reintegro en Cajero Propio” se gastaron 1.185.780€, lo que supone el 10,23% de la cantidad gastada con las tarjetas Black, siendo como se ha indicado anteriormente el segundo gasto más importante.
- El 36,09% de la cantidad total sacada del cajero (428.000€) se realizó en operaciones de entre 500 y 600€.
- El 48,58% de la cantidad total sacada del cajero (576.020€) se realizó en operaciones de entre 600 y 700 €.
- Estos dos tipos de operaciones suponen el 84,67% de la cantidad total gastada y 77,41% de las operaciones.

DISPOSICIÓN EFECTIVO OFICINA

Para las operaciones de “Disposición efectivo oficina” calculamos los siguientes estadísticos:

- La cantidad máxima sacada es de 13.549 €
- La cantidad media sacada es de 1.397,26€
- La cantidad mínima sacada es de 40€

Si tramificamos la cantidad gastada en intervalos de 250€:

	Cantidad (€)	%Cantidad	Frecuencia	%Frecuencia
<250 €	1.708	0,20%	13	2,17%
Entre 250 y 500 €	75.582	9,05%	248	41,47%
Entre 500 y 750€	62.424	7,47%	105	17,56%
Entre 750 y 1000 €	6.000	0,72%	7	1,17%
Entre 1000 y 1250 €	29.041	3,48%	27	4,52%
>1250 €	660.805	79,09%	198	33,11%
Total general	835.561	100%	598	100%

Tabla 35. Análisis de la operación Disposición Efectivo Oficina

En la Tabla 35 podemos ver como datos más relevantes:

- En concepto de “Disposición efectivo oficina” se gastaron 835.561€, lo que supone el 7,21% de la cantidad gastada con las tarjetas Black, siendo como se ha indicado anteriormente el tercer gasto más importante.
- El 79,09% de la cantidad dispuesta en cajeros (660.805 €) se sacó en operaciones de más de 1.250 €, suponiendo el 33,11% de las operaciones (segunda operación más frecuente).
- Las disposiciones que se realizaron con más frecuencia fueron de cantidades entre 250 y 500 €, suponiendo el 41,47% de las operaciones y el 9,05% de las cantidades dispuestas (75.582€).

REINTEGRO EN CAJERO AJENO NACIONAL

Para las operaciones de “Reintegro en cajero ajeno nacional” calculamos los siguientes estadísticos:

- La cantidad máxima sacada es de 610,11 €
- La cantidad media sacada es de 318,85€
- La cantidad mínima sacada es de 6,92€

Si tramificamos la cantidad gastada en intervalos de 100€:

	Cantidad	% Cantidad	Frecuencia	% Frecuencia
<100€	625	1,46%	13	9,70%
Entre 100 y 200 €	3.154	7,38%	23	17,16%
Entre 200 y 300 €	4.002	9,37%	17	12,69%
Entre 300 y 400 €	11.649	27,26%	36	26,87%
Entre 400 y 500 €	6.800	15,92%	15	11,19%
>500 €	16.495	38,61%	30	22,39%
Total general	42.726	100%	134	100%

Tabla 36. Análisis de la operación Cajero Ajeno Nacional

En la Tabla 36 podemos ver como datos más relevantes:

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

- En concepto de “Reintegro en cajero ajeno nacional” se gastaron 42.726€, lo que supone el 0,37% de la cantidad gastada con las tarjetas Black, siendo como se ha indicado anteriormente el cuarto gasto más importante
- La cantidad que con más frecuencia se sacaba del cajero fue de entre 300 y 400 €, seguida de cantidades superiores a 500€.

Las operaciones realizadas por estos dos rangos de cantidades, suponen el 65,87% del gasto total (28.144€)

CONSIDERANDO LA VARIABLE NOMBRE

Si mostramos la tabla de frecuencias de esta variable (para conocer el número de veces que cada consejero hizo de la tarjeta) y el gasto realizado en cada tipo de operación (para conocer la cantidad gastada por cada uno de ellos) obtenemos:

	NOMBRE			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
SÁNCHEZ BARCOJ, ILDEFONSO JOSÉ	2.743	3,63%	572.187,00	4,94%
MORAL SANTÍN, JOSÉ ANTONIO	1.215	1,61%	447.769,73	3,86%
MORADO IGLESIAS, RICARDO	265	0,35%	443.999,19	3,83%
AMAT ROCA, MATÍAS	1.036	1,37%	429.272,16	3,70%
BLESA DE LA PARRA, MIGUEL	1.565	2,07%	423.068,49	3,65%
FERRAZ RICARTE, RAMÓN	825	1,09%	389.450,35	3,36%
PÉREZ CLAVER, MARIANO	971	1,29%	354.171,00	3,06%

Tabla 37. Análisis de la variable Nombre

En la Tabla 37 podemos ver como datos más relevantes:

- Ildefonso José Sánchez Barcoj gastó 572.187€ (casi un 5% del total), el mayor gasto de todos los usuarios de las tarjetas.
- Por otro lado, José María García Alonso y Francisco Servando Verdú Pons fueron los consejeros que menos veces utilizaron la tarjeta (2 veces cada uno de ellos), sumando entre los dos un total de 251,30€.

CONSIDERANDO LA VARIABLE HORA

Obtenemos la tabla de frecuencias y gastos realizados con la variable “Hora”

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

	HORA			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
15	18.684	24,75%	2.271.977,40	19,61%
16	13.856	18,35%	1.943.748,28	16,77%
17	5.293	7,01%	899.934,19	7,77%
14	6.278	8,32%	895.305,13	7,73%
19	3.967	5,25%	635.593,41	5,48%
12	2.727	3,61%	586.476,41	5,06%
13	3.205	4,24%	560.650,50	4,84%
11	2.373	3,14%	532.814,48	4,60%
18	3.207	4,25%	510.208,84	4,40%
10	1.882	2,49%	454.842,68	3,92%
23	2.639	3,50%	423.227,88	3,65%
0	2.417	3,20%	409.567,87	3,53%
9	1.337	1,77%	358.869,93	3,10%
20	2.361	3,13%	335.735,26	2,90%
22	1.652	2,19%	207.674,68	1,79%
21	1.477	1,96%	189.406,32	1,63%
8	704	0,93%	152.023,16	1,31%
1	876	1,16%	129.643,83	1,12%
7	249	0,33%	40.180,88	0,35%
2	157	0,21%	31.265,22	0,27%
3	62	0,08%	5.723,56	0,05%
4	34	0,05%	5.721,50	0,05%
6	38	0,05%	4.640,26	0,04%
5	27	0,04%	3.397,41	0,03%
TOTAL	75.502	100%	11.588.629,08	100%

Tabla 38. Análisis de la variable Hora

- El gasto de las tarjetas es mayor entre las 15:00 y 16:00 de la tarde, suponiendo un 36,38% del gasto total y un 43,10% de la frecuencia de utilización
- Entre las 0:00 y las 6:00 a.m. se realizó un 5% del gasto total.
- Aunque el gasto realizado entre las 0:00 y las 6:00 a.m. no es elevado (sólo el 5% del gasto total), llama la atención por el momento del día que es, ya que las tarjetas Black estaban supuestamente destinadas a gastos de representación por temas laborales.

Vemos a continuación en qué se gastaron ese 5%.

OPERACIÓN	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
COMPRA	3.542	98,09%	571.503	96,87%
REINTEGRO EN CAJERO PROPIO	37	1,02%	19.360	3,28%
REINTEGRO EN CAJERO AJENO NACIONAL	1	0,03%	140	0,02%
CUOTA ANUAL TARJETA (BMP)	7	0,19%	132	0,02%
RECARGA TELEFONO GSM - CARGA EN F	1	0,03%	100	0,02%
DEVOLUCION CUOTA POR BMP	3	0,08%	-132	-0,02%
ANUL.COMPRAS, MARCAR/DESMARCAR FOTO	4	0,11%	-329	-0,06%
DEVOLUCION COMPRA (ABONO Y CARGO)	16	0,44%	-814	-0,14%
Total general	3.611	100%	589.960	100%

Tabla 39. Análisis de los gastos entre las 0:00 y 6:00

Vemos que la operación "Compra" fue la más realizada, seguida por "Reintegro en Cajero Propio"

Vemos en qué comercios se realizaron las **compras**

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

TIPO DE GASTO	Frecuencia	% Frecuencia	Cantidad(€)	%Cantidad
RESTAURANTES RESTO	1.133	32,01%	186.112	32,57%
EL CORTE INGLES	532	15,03%	112.703	19,72%
HIPERCOR SUPERMERCADOS EL CORTE INGLES	358	10,12%	68.580	12,00%
RESTO	1.516	42,84%	204.108	35,71%
Total	3.539	100%	571.503	100%

Tabla 40. Análisis de los gastos realizados en Compra entre las 0:00 y 6:00

CONSIDERANDO LA VARIABLE DÍA DE LA SEMANA

Obtenemos la tabla de frecuencias y gastos realizados con la variable “Día de la semana”

	DÍA DE LA SEMANA			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
VIERNES	11.887	15,74%	1.945.376,07	16,79%
MARTES	12.958	17,16%	1.868.641,26	16,12%
LUNES	10.562	13,99%	1.740.729,71	15,02%
MIÉRCOLES	11.185	14,81%	1.737.420,13	14,99%
JUEVES	10.800	14,30%	1.724.726,80	14,88%
SABADO	11.041	14,62%	1.627.486,79	14,04%
DOMINGO	7.069	9,36%	944.248,32	8,15%
TOTAL	75.502	100%	11.588.629,08	100%

Tabla 41. Análisis de la variable Día de la Semana

En la Tabla 41 podemos ver como datos más relevantes:

- El día de la semana que menor uso de las tarjetas se hizo fue los “Domingos”, con una frecuencia del 9,36% y suponiendo el 8,15% del gasto total
- El día que mayor uso se daba a las tarjetas eran los “Martes” con el 17,2%, seguido de los “Viernes” con un 15,7%. Entre estos dos días, el gasto realizado ascendía al 32,91% del total

CONSIDERANDO LA VARIABLE DÍA DEL MES

Si vemos la tabla de frecuencias, podemos ver cómo el uso y gasto de las tarjetas es prácticamente el mismo en cualquier día del mes.

CONSIDERANDO LA VARIABLE MES

Obtenemos la tabla de frecuencias y gastos realizados con la variable “Mes”

	MES			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
DICIEMBRE	6.801	9,01%	1.382.520,59	11,93%
JULIO	6.888	9,12%	1.042.555,13	9,00%
AGOSTO	6.421	8,50%	1.028.799,91	8,88%
MARZO	6.806	9,01%	1.021.958,96	8,82%
NOVIEMBRE	6.174	8,18%	974.238,43	8,41%
ABRIL	6.305	8,35%	956.297,00	8,25%
ENERO	6.095	8,07%	932.975,26	8,05%
MAYO	6.281	8,32%	908.522,28	7,84%
OCTUBRE	6.406	8,48%	907.934,48	7,83%
JUNIO	6.093	8,07%	869.884,30	7,51%
SEPTIEMBRE	6.168	8,17%	835.432,17	7,21%
FEBRERO	5.064	6,71%	727.510,57	6,28%
TOTAL	75.502	100%	11.588.629,08	100%

Tabla 42. Análisis de la variable Mes

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

En la Tabla 42 podemos ver como datos más relevantes:

- Los meses que mayor gasto se hizo fueron en Diciembre, con un total de 1.382.520,59€ (un 11,93% del total)
- Los meses que menor gasto se hizo fueron en Febrero, suponiendo un 6,28% del total.
- En Diciembre, Julio, Agosto y Marzo, el gasto realizado en cada uno de ellos fue mayor a 1.000.000€, siendo el resto de meses inferior a ese importe

CONSIDERANDO LA VARIABLE AÑO

Obtenemos la tabla de frecuencias y gastos realizados para la variable “Año”

	AÑO			
	Frecuencia	% Frecuencia	Cantidad (€)	% Cantidad
2008	8.857	11,73%	1.503.280,41	12,97%
2007	8.883	11,77%	1.485.122,64	12,82%
2010	9.557	12,66%	1.412.485,59	12,19%
2009	8.707	11,53%	1.395.278,90	12,04%
2006	8.737	11,57%	1.373.352,27	11,85%
2005	8.198	10,86%	1.187.475,07	10,25%
2011	7.947	10,53%	1.130.331,85	9,75%
2003	7.208	9,55%	1.068.883,34	9,22%
2004	7.212	9,55%	993.593,83	8,57%
2012	196	0,26%	38.825,18	0,34%
TOTAL	75.502	100%	11.588.629,08	100%

Tabla 43. Análisis de la variable Año

En la Tabla 43 podemos ver como datos más relevantes:

- El 50% de los gastos totales de las tarjetas se realizaron entre el año 2007 y 2010, siendo 2008 el año en que más gasto se realizó (un 12,97%) y 2010 el año que más veces se utilizaron las tarjetas.
- El 2012 fue el año que menos uso se dio a las tarjetas. En este caso hay que considerar también que los datos que se poseen son hasta Agosto de ese año.

Asimismo, se ha realizado un estudio del gasto y frecuencia de uso de las tarjetas por cada uno de los consejeros a través de distintos gráficos “Small Multiples” donde se representa con la misma escala y ejes, el gasto realizado por cada uno de los 82 consejeros en las diferentes variables, así como un estudio numérico de estos gastos.

El objetivo de este estudio ha sido el de detectar posibles usos diferentes a los establecidos para las tarjetas, sin embargo no se ha podido extraer ninguna conclusión más allá de una mera descripción de a qué destinaba cada consejero los gastos que realizaba y la frecuencia con que usaba la tarjeta en estos gastos, que no resulta de interés práctico para el presente estudio.

Se incluyen a continuación algunos de los gráficos más significativos. En todas estas representaciones, existe un gráfico por cada consejero, teniendo en el eje de las “y” el importe gastado y en el eje de las “x” los valores posibles de la variable que se está estudiando.

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

En la Figura 6 vemos como dato más destacable que los individuos 51 (Ricardo Morado Iglesias), 52 (José Antonio Moral Santín), 34 (Ramón Ferraz Ricarte), 77 (Enrique de la Torre Martínez) y 73 (Ildefonso José Sánchez Barcoj) son los que más gastaron en “Cajeros”. Este dato lo obtenemos también en el análisis de correspondencias cuyos resultados se muestran en el apartado 4.2.2.

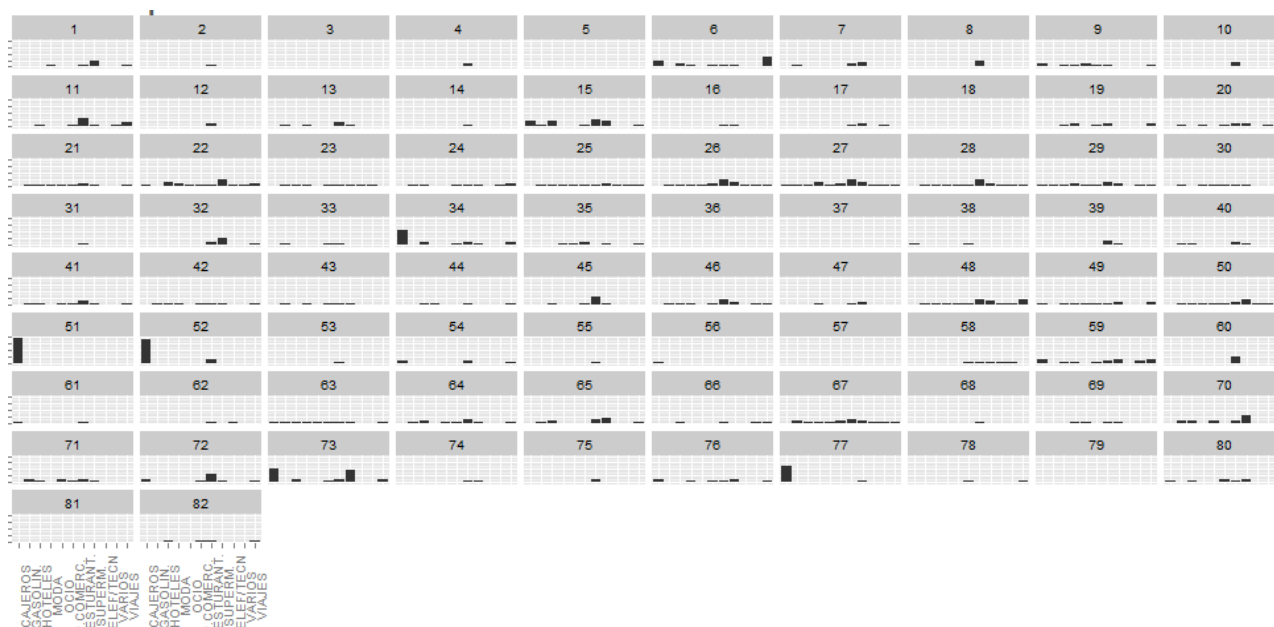


Figura 6. “Small Multiples” con la variable categoría

En la Figura 7 vemos por ejemplo que la observación 51 (Ricardo Morado Iglesias), realizó gastos mayores en los primeros días del mes que en los últimos; o que la observación 34 (Ramón Ferraz Ricarte) realizaba el mayor gasto en los días 20 de cada mes.

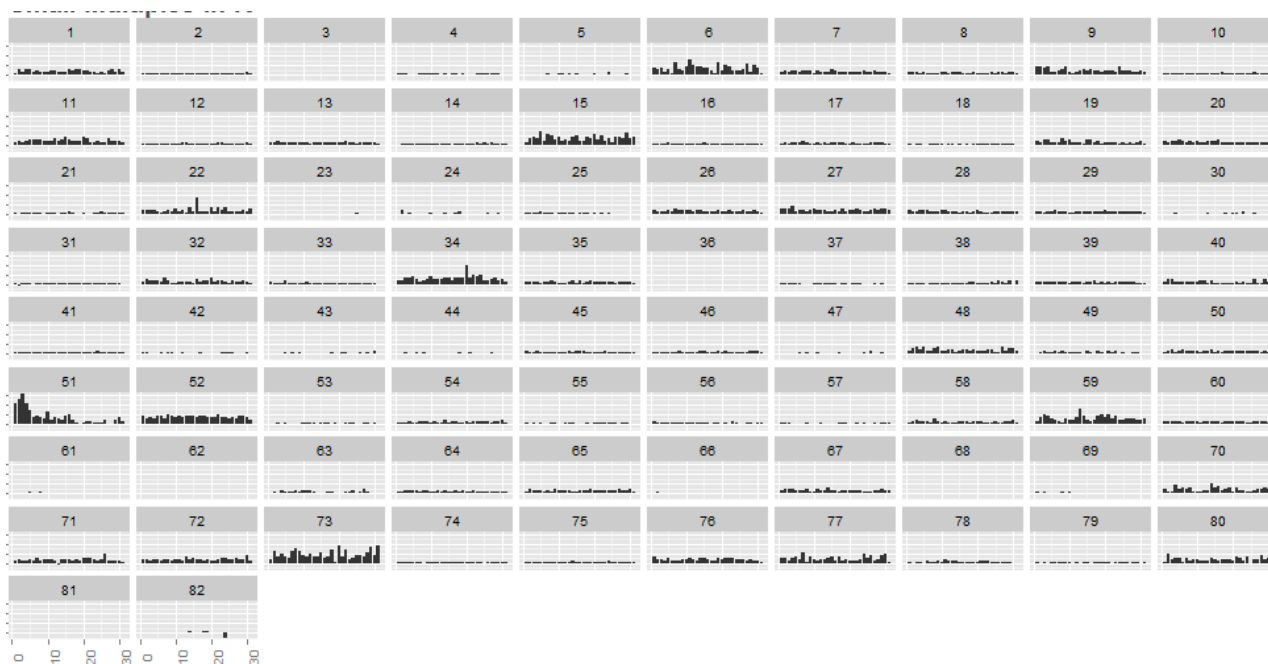


Figura 7: “Small Multiples” con la variable día del mes

En la Figura 8 vemos por ejemplo que la observación 70 (Estanislao Rodríguez-Ponga Salamanca) tiene un gasto similar a lo largo de todos los meses del año a excepción del mes de Diciembre donde el gasto se dispara (llegando a triplicarse). Lo mismo ocurre con el consejero 34 (Ramón Ferraz Ricarte), en el cual se llega a duplicar en Diciembre el gasto medio del resto de meses.

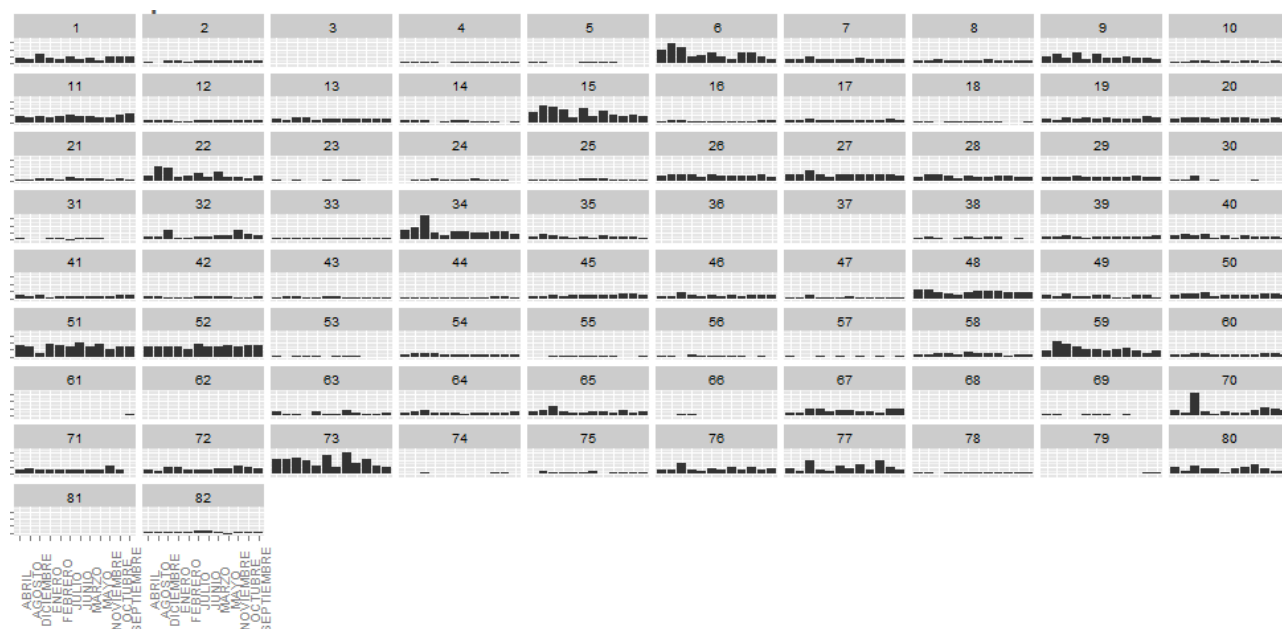


Figura 8: “Small Multiples” con la variable mes

4.2.2. Resultados del análisis de correspondencias

Análisis correspondencias simples entre nombre de consejero y tipo de operación

Representando gráficamente el análisis realizado y descrito en el apartado 3.2.4.1.

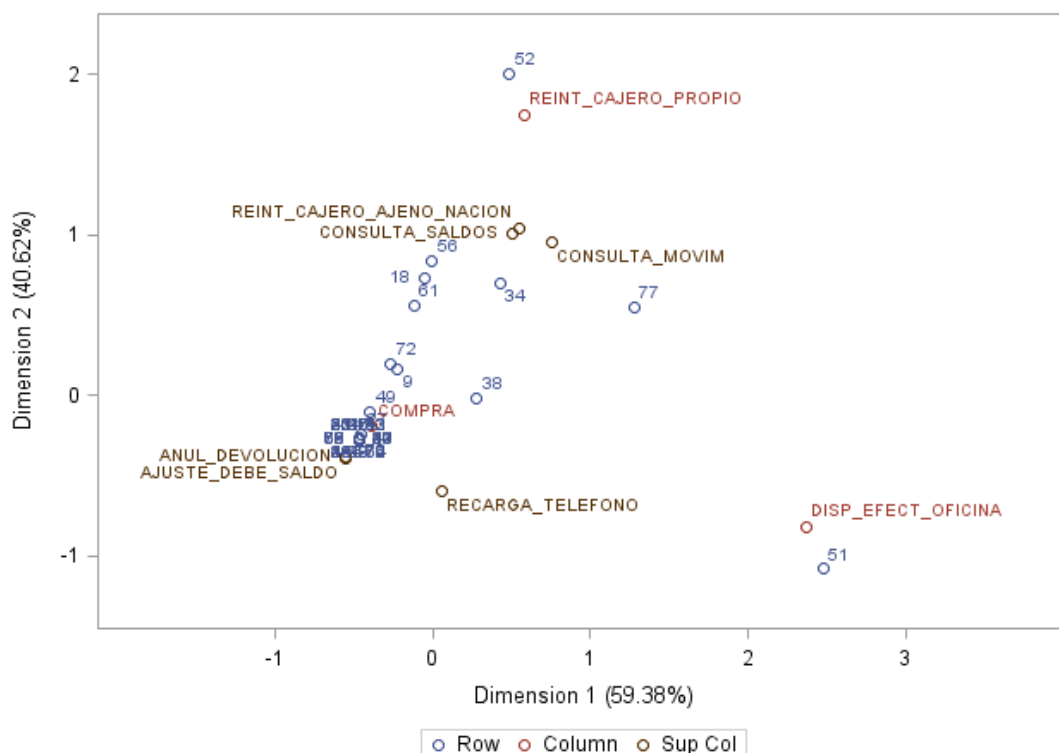


Figura 9: Análisis correspondencias simples entre nombre de consejero y tipo de operación

Podemos obtener las siguientes conclusiones de la Figura 9:

- La mayor parte de los consejeros centran el gasto realizado en la operación “Compra”
- El consejero Ricardo Morado Iglesias (51) se caracteriza por ser el que mayor gasto hizo en “Disposición de efectivo en Oficinas”.
- El consejero José Antonio Moral Santín (52) se caracteriza por ser el que mayor gasto hizo en “Reintegro en cajero Propio”.

Análisis correspondencias simples entre nombre de consejero y categoría de gasto

Si representamos gráficamente el análisis realizado y descrito en el apartado 3.2.4.1.

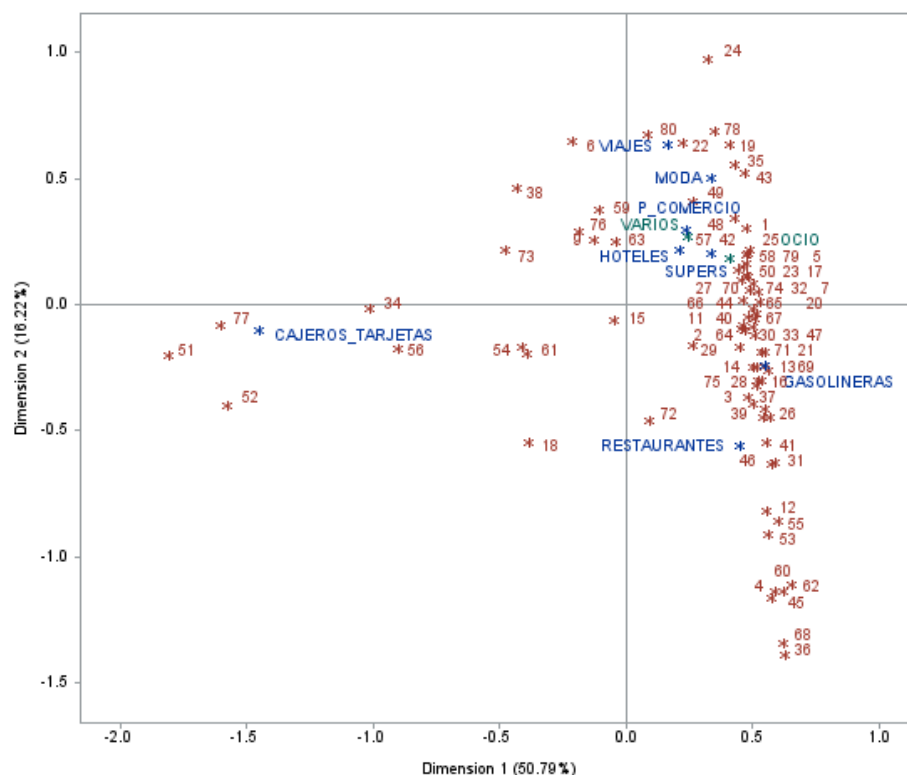


Figura 10: Análisis correspondencias simples entre nombre de consejero y categoría de gasto

Podemos obtener las siguientes conclusiones de la Figura 10:

- Los consejeros 77 (Enrique de la Torre Martínez), 51 (Ricardo Morado Iglesias), 52 (José Antonio Moran Santín), 34 (Ramón Ferraz Ricarte) y 56 (Domingo Navalmoral Sánchez) fueron los que mayor gasto realizaron en **cajeros y con las tarjetas**
- Los consejeros 80 (Carlos Vela García), 22 (Carmen Contreras Gómez), 78 (Rafael Eduardo Torres Posada), 19 (María Carmen Cafranga Cavestany), 35 (Luis Gabarda Durán), 43 (Juan Emilio Iranzo Martín) y 49 (Ramón Martínez Vilches) son los que más gastaron en **viajes y moda**.
- Los consejeros 41 (Jorge Gómez Moreno), 46 (Gonzalo Martín Pascual), 31 (José María Fernández del Río Fernández) fueron los que más gastaron en **restaurantes**.

Análisis correspondencias simples entre nombre de consejero y tramo horario

Representando gráficamente el análisis realizado y descrito en el apartado 3.2.4.1. podemos obtener las siguientes conclusiones de la Figura 11:

- Los consejeros 47 (José Ricardo Martínez Castro), 49 (Ramón Martínez Vilches) y 8 (José María Arteta Vico) fueron los que mayor gasto realizaron por la noche.
- La mayor parte del gasto se realizó en el tramo de la tarde.
- Los consejeros 18 (José Caballero Domínguez) y 24 (Miguel Corsini Freese) son los que más gastaron por la mañana.

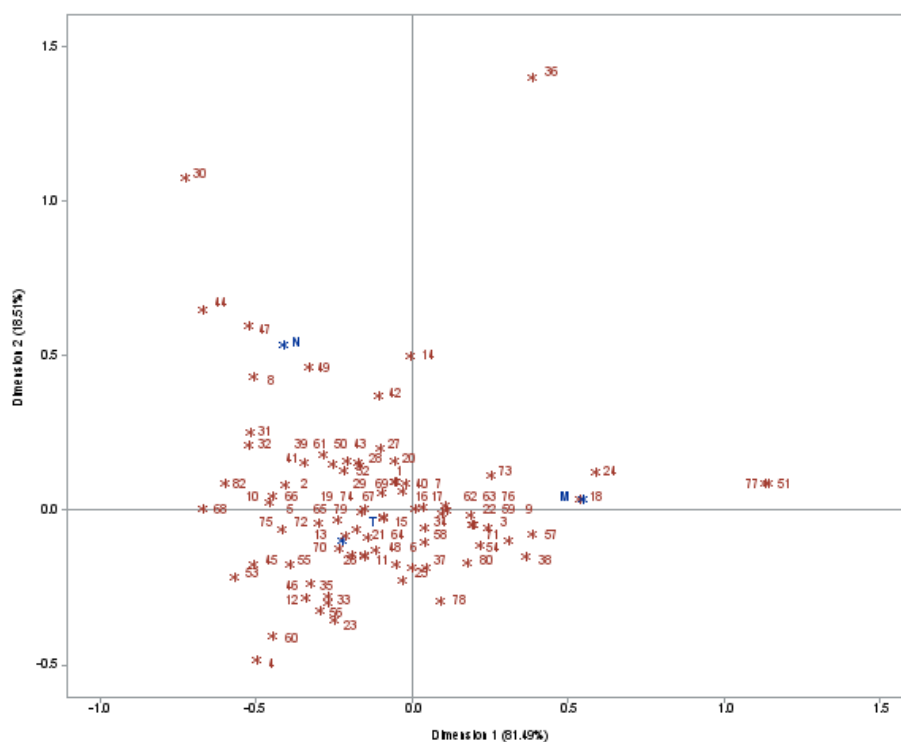


Figura 11: Análisis correspondencias simples entre nombre de consejero y tramo horario

Análisis correspondencias simples entre nombre de consejero y día de la semana

Representando gráficamente el análisis realizado y descrito en el apartado 3.2.4.1.

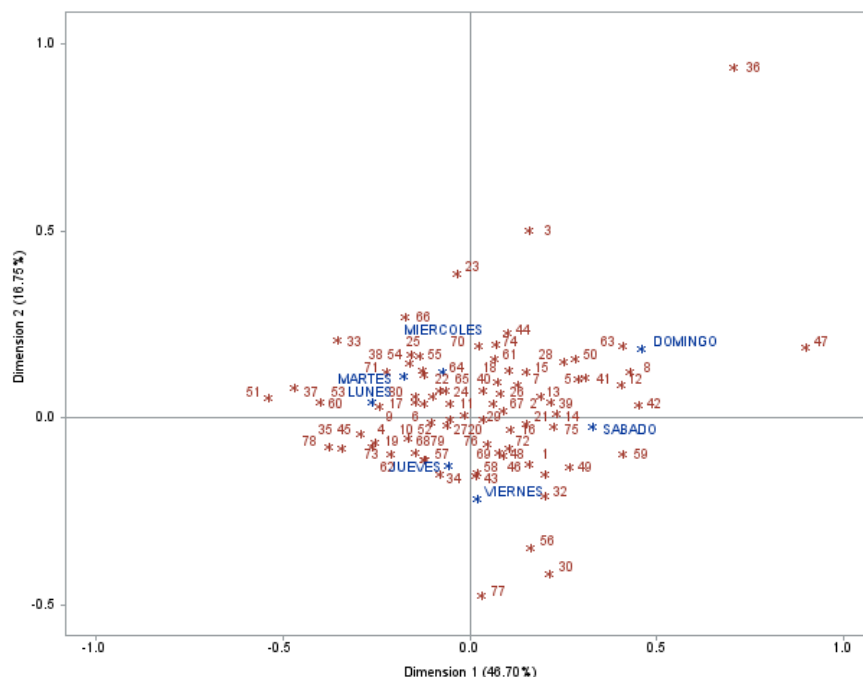


Figura 12: Análisis correspondencias simples entre nombre de consejero y día de la semana

A la vista de la Figura 12 podemos obtener las siguientes conclusiones:

- Los consejeros 8 (José María Arteta Vico), 63 (Rodrigo de Rato Figaredo) y 12 (Pedro Bedia Pérez) fueron los que mayor gasto realizaron en Domingo

4.2.3. Resultados del análisis clúster jerárquico

Tras el modelado mediante clusterización jerárquica, empleando distancia euclídea y algoritmo de Ward obtenemos 4 clúster con las siguientes observaciones cada uno de ellos:

- **CLUSTER 1 (31 observaciones):** Álvarez Gayol, María Enedina; Acero Benedicto, José Luis; Acosta Cubero, José; Bedia Pérez, Pedro; Blasco Bosqued, Luis; Caballero Domínguez, José; Cerón Escudero, Cándido; Contreras Gómez, José Carlos; Corsini Freese, Miguel; Fernández Álvarez, Arturo Luis; García Alonso, José María; García Pontes, Joaquín; Gutiérrez Moliner, Beltrán; Iranzo Martín, Juan Emilio; López Madrid, Francisco Javier; Martínez Castro, José Ricardo; Martínez Vilches, Ramón; Moreno Flores, Gabriel María; Muñiz de las Cuevas, Miguel; Navalmoral Sánchez, Domingo; Nieto Antolinos, José; Pradillo Moreno de la Santa, Rafael; Río García de Sola, Ignacio del; Rabago Juan Aracil, Jorge; Rizaldos González, Ángel; Rodríguez González, Manuel José; Sánchez Carlos, Santiago Javier; Serrano Antón, Fernando; Varela Díaz, Ignacio; Verdú Pons, Francisco Servando; Zapatero Gómez, Virgilio.
- **CLUSTER 2 (20 observaciones):** Abejón Resa, Miguel Ángel; Azcona Olondriz, Juan José; Buenaventura Zabala, Jose Maria; Bugidos Garay, Pedro; Cafranga Cavestany, María Carmen; Couceiro Ojeda, Alejandro; Fernández del Río Fernández, José María; Fernández Norriella, José Manuel; Fernández Yruegas Moro, Darío; Gómez Castañeda, Juan; Gómez del pulgar Perales, Ángel Eugenio; Gómez Moreno, Jorge; Gabarda Durán, Luis; Gil García, María Elena; Marcos Guerrero, Guillermo Ricardo;

Moure Bourio, Francisco José; Pérez Fernández, Francisco José; Rato Figaredo, Rodrigo de; Torres Posada, Rafael Eduardo; Vela García, Carlos.

- **CLUSTER 3 (23 observaciones):** Abejas Juarez, Pablo; Araujo Serrano, Miguel Ángel; Arteta Vico, José María; Astorqui Portera, Juan Manuel; Baquero Noriega, Francisco; Benito Valenciano, Rodolfo; Cámara Eguinoa, Antonio; Contreras Gómez, Carmen; Cruz Orive, Rubén; de la Merced Monge, María Mercedes; de Navasques Cobián, Ignacio; Espinar Gallego, Ramón; Martín Pascual, Gonzalo; Martínez Martínez, Carlos María; Miguel Sánchez, Javier de; Pedroche Nieto, Jesús; Recarte García Andrade, Alberto; Rey de Viñas Sánchez-Majestad, Antonio; Riva Ámez, José María de la; Rodríguez-Ponga Salamanca, Estanislao; Romero de Tejada y Picatoste, Ricardo; Romero Lázaro, Antonio; Spottorno Díaz Caro, Rafael.
- **CLUSTER 4 (7 observaciones):** Amat Roca, Matías; Blesa de la Parra, Miguel; Ferraz Ricarte, Ramón; Morado iglesias, Ricardo; Moral Santín, José Antonio; Pérez Claver, Mariano; Torre Martínez, Enrique de la.

Por otra parte Ildefonso Sánchez Barcoj que lo estudiamos por separado.

Si vemos las características que definen a cada uno de estos clúster, los podemos caracterizar de la siguiente forma:

- **CLUSTER 1:** se trata del grupo con menor gasto y menor frecuencia de uso de las tarjetas en casi todas las variables de estudio (a excepción de la frecuencia de uso de 2011 y 2012).
- **CLUSTER 2:** este grupo destaca por la frecuencia de gasto e importe durante el año 2012.
- **CLUSTER 3:** se trata del grupo con mayor gasto y mayor frecuencia de uso de las tarjetas en comercios, telefonía, restaurantes, supermercados, moda, gasolineras y ocio. Fueron los que más emplearon la tarjeta y más gastaron en 2011, y en el resto de años fueron los que más emplearon la tarjeta, aunque no fueran los que más gasto hacían. Tanto para los días de la semana, meses y franjas horarias, eran los que empleaban la tarjeta con mayor frecuencia, sin embargo no eran los que más gastaban.
- **CLUSTER 4:** se trata del grupo con mayor gasto en hoteles, viajes y cajeros. Fueron los que más gastaron entre 2004 y 2010, y los que menos en 2011 y 2012, sin embargo durante los años que más gastaron no fueron los que más veces emplearon las tarjetas, por lo que las compras que realizaban eran de importes más elevados que la media de los otros consejeros. Esto mismo ocurre con los días de la semana, meses y franjas horarias, siendo los que mayor importe medio gastaron pero no los que empleaban la tarjeta con más frecuencia.
- Por último tenemos a **Ildefonso José Sánchez Barcoj**, cuyo gasto y frecuencia de uso de la tarjeta en comercios (5,04; 6,10) y supermercados (8,27; 6,07) es mucho mayor que la media de los otros clúster (3,93; 4,25), así como el uso de la tarjeta en ocio y cajeros (8,63 frente a 3,74 y 6,94 frente a 5,55).

El gasto realizado y la frecuencia de uso durante el año 2012 (9,24; 9,78) es también muy superior a la media de los otros clúster (3,57; 3,56). Asimismo, el importe gastado en los meses de Marzo, Abril y

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

Julio (8,62 frente a 4,92; 7,21 frente a 5,32; 7,81 frente a 5,20) y la frecuencia de uso durante el mes de Agosto (7,94 frente a 4,17) es también muy superior.

Se muestra en la Tabla 44 las medias de cada variable de estudio empleadas para cada clúster.

cluster	1	2	3	4	cluster	1	2	3	4	cluster	1	2	3	4	cluster	1	2	3	4
avg (Imp_pComerci...)	2,58	3,42	3,93	3,89	avg (Imp_2003)	2,78	3,31	3,36	5,11	avg (Imp_Febrero)	2,49	3,24	3,79	5,13	avg (Imp_Lunes)	2,57	3,16	3,74	4,93
avg (Frec_pComerci...)	2,66	3,21	4,25	2,93	avg (Frec_2003)	2,79	3,42	3,71	3,53	avg (Frec_Febrero)	2,46	3,25	4,48	3,38	avg (Frec_Lunes)	2,43	3,29	4,48	3,21
avg (Imp_telef)	2,84	3,21	4,85	3,22	avg (Impn_2004)	2,58	3,41	3,59	5,38	avg (Impn_Marzo)	2,54	3,18	3,71	4,92	avg (Impn_Martes)	2,48	3,21	3,82	5,85
avg (Frec_telef)	2,63	3,48	4,86	3,29	avg (Frec_2004)	2,48	3,54	4,87	3,51	avg (Frec_Marzo)	2,47	3,21	4,46	3,17	avg (Frec_Martes)	2,43	3,32	4,37	3,31
avg (Imp_restaur)	2,64	3,35	4,18	3,58	avg (Imp_2005)	2,52	3,34	3,68	5,33	avg (Imp_Abril)	2,56	3,18	3,78	5,32	avg (Imp_Miercoles)	2,54	3,12	3,73	5,44
avg (Frec_restaur)	2,78	3,33	4,17	3,86	avg (Frec_2005)	2,49	3,43	4,17	3,43	avg (Frec_Abril)	2,49	3,16	4,42	3,26	avg (Frec_Miercoles)	2,51	3,23	4,42	3,22
avg (Impn_varios)	2,82	3,88	3,82	4,29	avg (Impn_2006)	2,49	3,33	3,64	5,49	avg (Imp_Mayo)	2,58	3,23	3,71	5,35	avg (Impn_Lueves)	2,55	3,16	3,65	5,33
avg (Frec_varios)	2,77	3,34	4,85	3,87	avg (Frec_2006)	2,48	3,48	4,32	3,45	avg (Frec_Mayo)	2,45	3,22	4,38	3,36	avg (Frec_Lueves)	2,51	3,23	4,39	3,24
avg (Impn_suvers)	2,75	3,13	3,91	3,55	avg (Impn_2007)	2,47	3,86	3,95	5,21	avg (Imp_Junio)	2,54	3,11	3,78	5,45	avg (Impn_Viernes)	2,55	3,84	3,78	5,62
avg (Frec_suvers)	2,75	3,11	4,22	2,93	avg (Frec_2007)	2,58	3,88	4,49	3,44	avg (Frec_Junio)	2,52	3,17	4,48	3,38	avg (Frec_Viernes)	2,47	3,21	4,41	3,29
avg (Impn_moda)	2,84	3,23	3,98	3,61	avg (Impn_2008)	2,57	2,85	4,82	5,19	avg (Imp_Julio)	2,59	2,98	3,78	5,28	avg (Impn_Sabado)	2,51	3,82	4,89	4,74
avg (Frec_moda)	2,91	3,26	3,93	3,83	avg (Frec_2008)	2,59	2,92	4,53	3,35	avg (Frec_Julio)	2,47	3,13	4,48	3,28	avg (Frec_Sabado)	2,55	3,81	4,46	3,25
avg (Imp_hotel)	2,73	3,18	3,69	4,68	avg (Imp_2009)	2,61	2,84	3,95	5,89	avg (Imp_Agosto)	2,65	3,82	3,66	5,46	avg (Imp_Domingo)	2,65	2,94	3,94	5,88
avg (Frec_hotel)	2,66	3,19	4,81	3,93	avg (Frec_2009)	2,62	2,89	4,58	3,37	avg (Frec_Agosto)	2,56	3,89	4,17	3,78	avg (Frec_Domingo)	2,61	2,95	4,43	3,43
avg (Imp_gasolin)	2,69	3,29	4,21	3,86	avg (Imp_2010)	2,95	3,86	3,66	4,15	avg (Imp_Septiembr...)	2,47	3,11	4,88	5,81	avg (Imp_Manana)	2,71	3,83	3,54	5,48
avg (Frec_gasolin)	2,66	3,38	4,26	2,96	avg (Frec_2010)	3,82	3,28	3,89	2,85	avg (Frec_Septiembr...)	2,47	3,28	4,49	3,13	avg (Frec_Manana)	2,54	3,89	4,37	3,39
avg (Imp_ocio)	2,98	3,17	4,82	3,88	avg (Imp_2011)	3,14	3,18	3,58	3,28	avg (Imp_Octubre)	2,44	3,15	3,99	5,89	avg (Imp_Tarde)	2,44	3,19	3,96	4,92
avg (Frec_ocio)	2,88	3,38	3,74	3,88	avg (Frec_2011)	3,16	3,19	3,65	2,83	avg (Frec_Octubre)	2,49	3,22	4,44	3,18	avg (Frec_Tarde)	2,46	3,24	4,44	3,22
avg (Imp_viajes)	2,89	3,29	3,45	4,52	avg (Imp_2012)	3,87	3,57	3,17	3,17	avg (Imp_Noviembr...)	2,51	3,16	3,83	4,98	avg (Imp_Noche)	2,72	3,13	3,82	4,22
avg (Frec_viajes)	2,66	3,13	4,31	3,18	avg (Frec_2012)	3,12	3,56	3,11	3,11	avg (Frec_Noviembr...)	2,51	3,21	4,47	3,15	avg (Frec_Noche)	2,78	3,13	4,18	3,25
avg (Imp_cajeros)	2,98	3,85	3,88	5,87	avg (Imp_Enero)	2,52	3,86	3,85	5,17	avg (Imp_Diciembre...)	2,61	3,84	3,81	5,12					
avg (Frec_cajeros)	2,98	3,85	3,13	5,55	avg (Frec_Enero)	2,48	3,24	4,39	3,26	avg (Frec_Diciembre...)	2,49	3,15	4,53	3,16					

Tabla 44: Medias de cada variable en cada clúster. Método jerárquico

4.2.4. Resultados del análisis clúster no jerárquico (k-means)

Tras el modelado mediante clusterización k-means obtenemos 4 clúster con las siguientes observaciones cada uno de ellos:

- **CLUSTER 1 (24 observaciones):** Abejas Juarez, Pablo; Arteta Vico, José María; Astorqui Portera, Juan Manuel; Baquero Noriega, Francisco; Benito Valenciano, Rodolfo; Contreras Gómez, Carmen; de la Merced Monge, María Mercedes; de Navasques Cobián, Ignacio; Fernández Nornieilla, José Manuel; Ferraz Ricarte, Ramón; Gómez Moreno, Jorge; Gabarda Durán, Luis; Martín Pascual, Gonzalo; Martínez Martínez, Carlos María; Miguel Sánchez, Javier de; Pedroche Nieto, Jesús; Recarte García Andrade, Alberto; Rey de Viñas Sánchez-Majestad, Antonio; Riva Ámez, José María de la; Rodríguez-Ponga

Salamanca, Estanislao; Romero Lázaro, Antonio; Spottorno Díaz Caro, Rafael; Torre Martínez, Enrique de la; Vela García, Carlos.

- **CLUSTER 2 (10 observaciones):** Amat Roca, Matías; Araujo Serrano, Miguel Ángel; Blesa de la Parra, Miguel; Cámara Eguinoa, Antonio; Cruz Orive, Rubén; Espinar Gallego, Ramón; Morado Iglesias, Ricardo; Moral Santín, José Antonio; Pérez Claver, Mariano; Romero de Tejada y Picatoste, Ricardo.
- **CLUSTER 3 (18 observaciones):** Abejón Resa, Miguel Ángel; Azcona Olondriz, Juan José; Bedia Pérez, Pedro; Buenaventura Zabala, José María; Bugidos Garay, Pedro; Cafranga Cavestany, María Carmen; Cerón Escudero, Cándido; Couceiro Ojeda, Alejandro; Fernández Yruegas Moro, Darío; Gómez Castañeda, Juan; Gómez del Pulgar Perales, Ángel Eugenio; Marcos Guerrero, Guillermo Ricardo; Martínez Vilches, Ramón; Moure Bourio, Francisco José; Pérez Fernández, Francisco José; Rato Figaredo, Rodrigo de; Serrano Antón, Fernando; Torres Posada, Rafael Eduardo.
- **CLUSTER 4 (29 observaciones):** Álvarez Gayol, María Enedina; Acero Benedicto, José Luis; Acosta Cubero, José; Blasco Bosqued, Luis; Caballero Domínguez, José; Contreras Gómez, José Carlos; Corsini Freese, Miguel; Fernández Álvarez, Arturo Luis; Fernández del Río Fernández, José María; García Alonso, José María; García Pontes, Joaquín; Gil García, María Elena; Gutiérrez Moliner, Beltrán; Iranzo Martín, Juan Emilio; López Madrid, Francisco Javier; Martínez Castro, José Ricardo; Moreno Flores, Gabriel María; Muñiz de las Cuevas, Miguel; Navalmoral Sánchez, Domingo; Nieto Antolinos, José; Pradillo Moreno de la Santa, Rafael; Río García de Sola, Ignacio del; Rabago Juan Aracil, Jorge; Rizaldos González, Ángel; Rodríguez González, Manuel José; Sánchez Carlos, Santiago Javier; Varela Díaz, Ignacio; Verdú Pons, Francisco Servando; Zapatero Gómez, Virgilio.

Si vemos las características que definen a cada uno de estos clúster, los podemos caracterizar de la siguiente forma:

- **CLUSTER 1:** se trata del grupo que más viaja y el segundo que más gasta en viaje, son los que más gastan en Ocio y los segundos con mayor frecuencia en este tipo de gasto. Además son los que más gastan en Moda y con más frecuencia, los que van a los supermercados con mayor frecuencia y los que más gastan en telefonía y con mayor frecuencia.
- **CLUSTER 2:** se trata del grupo con mayor gasto y mayor frecuencia de uso de las tarjetas en casi todas las variables de estudio.
- **CLUSTER 3:** este grupo no destaca, ni por exceso ni por defecto, en ningún tipo de gasto. Únicamente se puede destacar de ellos que fueron los que menos gastaron y menor uso hicieron de las tarjetas en el año 2011 y que al año siguiente, en 2012, pasaron a ser los que más gastaron y mayor uso hicieron de las mismas.
- **CLUSTER 4:** se trata del grupo con menor gasto y menor frecuencia de uso de las tarjetas en casi todas las variables de estudio.

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

- **Ildefonso Sánchez Barcoj**: este consejero, que fue el que mayor gasto realizó con las tarjetas, se caracteriza por tener gastos y frecuencias de uso de la tarjeta mucho más elevadas que el resto de consejeros en casi todas las variables. Estas diferencias son más visibles en el importe gastado en supermercados (8,27 frente a 3,87), la frecuencia de gasto en ocio (8,63 frente a 3,87), la frecuencia de uso en cajeros (6,94 frente a 4,25), la frecuencia y gasto en el año 2012 (9,78 frente a 3,35 y 9,24 frente a 3,46), el importe gastado en los meses de Marzo (8,62 frente a 2,55), Abril (7,21 frente a 4,67), Julio (7,81 frente a 4,55) y Agosto (7,94 frente a 4,40)...

Se muestra en la Tabla 45 las medias de cada variable de estudio empleadas para cada clúster.

clus	1	2	3	4
avg (Imp_pComerci...)	3,94	4,04	3,04	2,62
avg (Frec_pComerci...)	3,75	4,28	3,03	2,67
avg (Imp_telef)	3,88	3,57	3,14	2,85
avg (Frec_telef)	3,93	3,78	3,26	2,64
avg (Imp_restaur)	3,85	3,92	3,47	2,54
avg (Frec_restaur)	3,69	4,23	3,39	2,61
avg (Imp_varios)	3,57	4,45	3,08	2,81
avg (Frec_varios)	3,74	3,93	3,09	2,85
avg (Imp_sinere)	3,87	3,80	2,95	2,71
avg (Frec_sinere)	3,66	4,38	3,00	2,73
avg (Imp_moda)	3,78	3,70	3,15	2,86
avg (Frec_moda)	3,71	3,70	3,11	2,93
avg (Imp_hotel)	3,77	4,03	3,19	2,67
avg (Frec_hotel)	3,80	4,13	3,31	2,55
avg (Imp_gasolin)	3,67	4,29	3,26	2,68
avg (Frec_gasolin)	3,70	4,21	3,31	2,65
avg (Imp_ocio)	3,78	3,67	3,10	2,88
avg (Frec_ocio)	3,45	3,87	3,17	2,89
avg (Imp_viajes)	3,64	3,93	3,18	2,84
avg (Frec_viajes)	4,07	3,90	2,97	2,63
avg (Imp_cajeros)	3,35	4,38	3,02	2,99
avg (Frec_cajeros)	3,36	4,25	3,02	2,98

clus	1	2	3	4
avg (Imp_2003)	3,46	4,42	2,99	2,91
avg (Frec_2003)	3,36	4,25	3,17	2,92
avg (Imp_2004)	3,65	4,64	3,13	2,59
avg (Frec_2004)	3,65	4,39	3,41	2,51
avg (Imp_2005)	3,76	4,43	3,14	2,54
avg (Frec_2005)	3,75	4,33	3,28	2,53
avg (Imp_2006)	3,78	4,62	3,13	2,50
avg (Frec_2006)	3,88	4,40	3,25	2,42
avg (Imp_2007)	4,08	4,40	2,95	2,42
avg (Frec_2007)	4,03	4,38	3,06	2,44
avg (Imp_2008)	3,97	4,43	3,00	2,47
avg (Frec_2008)	3,90	4,49	3,10	2,47
avg (Imp_2009)	3,91	4,37	3,08	2,45
avg (Frec_2009)	3,90	4,46	3,16	2,45
avg (Imp_2010)	3,57	4,24	3,10	2,82
avg (Frec_2010)	3,64	3,94	3,09	2,92
avg (Imp_2011)	3,44	3,65	3,03	3,13
avg (Frec_2011)	3,41	3,87	3,00	3,15
avg (Imp_2012)	3,29	3,17	3,46	3,06
avg (Frec_2012)	3,31	3,11	3,35	3,12
avg (Imp_Enero)	3,77	4,74	3,03	2,48
avg (Frec_Enero)	3,79	4,63	3,21	2,43

clus	1	2	3	4
avg (Imp_Febrero)	3,71	4,79	3,13	2,47
avg (Frec_Febrero)	3,77	4,76	3,16	2,44
avg (Imp_Marzo)	3,79	4,32	3,02	2,55
avg (Frec_Marzo)	3,90	4,55	3,08	2,45
avg (Imp_Abril)	3,75	4,67	2,98	2,54
avg (Frec_Abril)	3,84	4,62	3,03	2,48
avg (Imp_Mayo)	3,77	4,74	3,04	2,48
avg (Frec_Mayo)	3,85	4,60	3,11	2,42
avg (Imp_Junio)	3,80	4,75	3,09	2,47
avg (Frec_Junio)	3,79	4,63	3,17	2,46
avg (Imp_Julio)	3,75	4,55	3,01	2,53
avg (Frec_Julio)	3,85	4,64	3,09	2,44
avg (Imp_Agosto)	3,65	4,79	2,96	2,63
avg (Frec_Agosto)	3,87	4,40	2,99	2,49
avg (Imp_Septiembr...)	3,98	4,47	3,08	2,41
avg (Frec_Septiembr...)	3,91	4,51	3,14	2,41
avg (Imp_Octubre)	4,00	4,53	3,13	2,35
avg (Frec_Octubre)	3,89	4,52	3,17	2,43
avg (Imp_Noviembr...)	3,96	4,35	3,01	2,46
avg (Frec_Noviembr...)	3,86	4,63	3,17	2,44
avg (Imp_Diciembre...)	4,07	4,05	2,93	2,56
avg (Frec_Diciembre...)	4,06	4,36	3,08	2,42

clus	1	2	3	4
avg (Imp_Lunes)	3,76	4,42	3,06	2,55
avg (Frec_Lunes)	3,87	4,47	3,19	2,39
avg (Imp_Martes)	3,82	4,62	3,06	2,47
avg (Frec_Martes)	3,79	4,65	3,20	2,42
avg (Imp_Miercoles)	3,67	4,87	3,04	2,54
avg (Frec_Miercoles)	3,80	4,67	3,15	2,48
avg (Imp_Jueves)	3,84	4,39	3,08	2,48
avg (Frec_Jueves)	3,80	4,68	3,16	2,45
avg (Imp_Viernes)	3,87	4,50	2,96	2,53
avg (Frec_Viernes)	3,88	4,56	3,11	2,42
avg (Imp_Sabado)	4,12	4,34	2,93	2,44
avg (Frec_Sabado)	3,97	4,35	2,98	2,49
avg (Imp_Domingo)	3,87	4,61	3,00	2,54
avg (Frec_Domingo)	3,95	4,46	2,96	2,51
avg (Imp_Manana)	3,65	4,49	2,96	2,71
avg (Frec_Manana)	3,80	4,64	2,96	2,54
avg (Imp_Tarde)	3,96	4,53	3,09	2,38
avg (Frec_Tarde)	3,90	4,55	3,19	2,40
avg (Imp_Noche)	3,79	4,07	3,09	2,65
avg (Frec_Noche)	3,75	4,23	3,11	2,65

Tabla 45: Medias de cada variable en cada clúster. Método k-means

4.2.5. Resultados del análisis clúster difuso (fuzzy clustering)

Tras el modelado mediante clusterización difusa obtenemos 4cluster con las siguientes observaciones cada uno de ellos:

- **CLUSTER 1 (16 observaciones):** Álvarez Gayol, María Enedina; Azcona Olondriz, Juan José; Bedia Pérez, Pedro; Buenaventura Zabala, José María; Cerón Escudero, Cándido; Couceiro Ojeda, Alejandro; Fernández Yruegas Moro, Darío; Gutiérrez Moliner, Beltrán; Marcos Guerrero, Guillermo Ricardo; Martínez Vilches, Ramón; Moure Bourio, Francisco José; Pérez Fernández, Francisco José; Rato Figaredo, Rodrigo de; Sánchez Carlos, Santiago Javier; Serrano Antón, Fernando; Torres Posada, Rafael Eduardo
- **CLUSTER 2 (12 observaciones):** Abejón Resa, Miguel Ángel; Bugidos Garay, Pedro; Cafranga Cavestany, María Carmen; Fernández Norniella, José Manuel; Gómez Castañeda, Juan; Gómez Moreno, Jorge; Gómez del Pulgar Perales, Ángel Eugenio; Gómez Moreno, Jorge; Gabarda Durán, Luis; Martín Pascual, Gonzalo; Morado Iglesias, Ricardo; Recarte García Andrade, Alberto; Torre Martínez, Enrique de la.
- **CLUSTER 3 (28 observaciones):** Abejas Juarez, Pablo; Amat Roca, Matías; Araujo Serrano, Miguel Ángel; Arteta Vico, José María; Astorqui Portera, Juan Manuel; Baquero Noriega, Francisco; Benito Valenciano, Rodolfo; Blesa de la Parra, Miguel; Cámara Eguinoa, Antonio; Contreras Gómez, Carmen; Cruz Orive, Rubén; de la Merced Monge, María Mercedes; de Navasques Cobián, Ignacio; Espinar Gallego, Ramón; Ferraz Ricarte, Ramón; Martínez Martínez, Carlos María; Miguel Sánchez, Javier de; Moral Santín, José Antonio; Pérez Claver, Mariano; Pedroche Nieto, Jesús; Rey de Viñas Sánchez-Majestad, Antonio; Riva Ámez, José María de la; Rodríguez-Ponga Salamanca, Estanislao; Romero de Tejada y Picatoste, Ricardo; Romero Lázaro, Antonio; Sánchez Barcoj, Ildefonso José; Spottorno Díaz Caro, Rafael; Vela García, Carlos.
- **CLUSTER 4 (26 observaciones):** Acero Benedicto, José Luis; Acosta Cubero, José; Blasco Bosqued, Luis; Caballero Domínguez, José; Contreras Gómez, José Carlos; Corsini Freese, Miguel; Fernández Álvarez, Arturo Luis; Fernández del Río Fernández, José María; García Alonso, José María; García Pontes, Joaquín; Gil García, María Elena; Iranzo Martín, Juan Emilio; López Madrid, Francisco Javier; Martínez Castro, José Ricardo; Moreno Flores, Gabriel María; Muñiz de las Cuevas, Miguel; Navalmoral Sánchez, Domingo; Nieto Antolinos, José; Pradillo Moreno de la Santa, Rafael; Río García de Sola, Ignacio del; Rabago Juan Aracil, Jorge; Rizaldos González, Ángel; Rodríguez González, Manuel José; Varela Díaz, Ignacio; Verdú Pons, Francisco Servando; Zapatero Gómez, Virgilio.

Si vemos las características que definen a cada uno de estos clúster, los podemos caracterizar:

- **CLUSTER 1:** se trata del grupo con menor gasto durante el año 2003 y con mayor gasto durante el año 2012.
- **CLUSTER 2:** este grupo únicamente destaca por ser el que mayor frecuencia de uso de las tarjetas tuvo durante el año 2012.
- **CLUSTER 3:** se trata del grupo con mayor gasto y mayor frecuencia de uso de las tarjetas en casi todas las variables de estudio a excepción del importe y frecuencia de uso durante el año 2012 donde se sitúan los terceros.

USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

- **CLUSTER 4:** se trata del grupo con menor gasto y menor frecuencia de uso de las tarjetas en todas las variables de estudio a excepción del importe gastado en 2013 donde se sitúan los segundos en cuanto a menor gasto.

Se muestra en la Tabla 46 las medias de cada variable de estudio empleadas para cada clúster.

Cluster	1	2	3	4
avg (Imp_pComerci...)	2,96	3,19	4,17	2,64
avg (Frec_pComerci...)	2,91	3,22	4,16	2,66
avg (Imp_telef)	3,14	3,04	3,95	2,83
avg (Frec_telef)	3,25	3,28	3,97	2,63
avg (Imp_restaur)	3,32	3,38	3,99	2,53
avg (Frec_restaur)	3,30	3,42	3,93	2,58
avg (Impn variac)	2,90	3,22	4,83	2,81
avg (Frec variac)	2,91	3,37	3,91	2,86
avg (Impn super)	2,87	3,38	4,89	2,69
avg (Frec super)	2,91	3,24	4,11	2,71
avg (Impn moda)	2,93	3,37	3,94	2,82
avg (Frec moda)	3,00	3,38	3,85	2,86
avg (Imp_hotel)	2,95	3,42	4,86	2,66
avg (Frec_hotel)	3,03	3,58	4,85	2,55
avg (Imp_gasolin)	3,08	3,31	4,83	2,66
avg (Frec_gasolin)	3,11	3,32	4,83	2,64
avg (Imp_ocio)	3,04	3,27	3,80	2,86
avg (Frec_ocio)	3,09	3,36	3,82	2,86
avg (Imp_viajes)	3,08	3,28	3,98	2,82
avg (Frec_viajes)	2,93	3,59	4,88	2,57
avg (Imp_cajeros)	3,03	3,71	3,58	2,99
avg (Frec_cajeros)	3,02	3,28	3,78	2,98

Cluster	1	2	3	4
avg (Imp_2003)	2,93	3,21	3,86	2,98
avg (Frec_2003)	3,08	3,03	3,84	2,99
avg (Impn 2004)	2,95	3,56	4,85	2,61
avg (Frec 2004)	3,14	3,49	4,84	2,52
avg (Imp_2005)	2,98	3,60	4,86	2,55
avg (Frec_2005)	3,07	3,37	4,12	2,54
avg (Impn 2006)	2,93	3,52	4,17	2,50
avg (Frec 2006)	3,00	3,41	4,25	2,43
avg (Impn 2007)	2,73	3,70	4,28	2,42
avg (Frec 2007)	2,85	3,45	4,31	2,44
avg (Impn 2008)	2,82	3,51	4,27	2,47
avg (Frec 2008)	2,90	3,31	4,31	2,47
avg (Imp_2009)	2,89	3,52	4,24	2,46
avg (Frec_2009)	2,92	3,31	4,31	2,45
avg (Imp_2010)	3,17	3,53	3,83	2,72
avg (Frec_2010)	3,26	3,48	3,75	2,77
avg (Imp_2011)	3,12	3,34	3,66	3,02
avg (Frec_2011)	3,21	3,37	3,62	3,00
avg (Imp_2012)	3,50	3,41	3,39	3,05
avg (Frec_2012)	3,37	3,51	3,35	3,12
avg (Imp_Enero)	2,83	3,56	4,23	2,47
avg (Frec_Enero)	2,98	3,39	4,28	2,42

Cluster	1	2	3	4
avg (Imp_Febrero)	2,93	3,56	4,17	2,48
avg (Frec_Febrero)	3,00	3,36	4,27	2,43
avg (Impn Marzo)	2,89	3,57	4,14	2,53
avg (Frec Marzo)	2,95	3,33	4,32	2,41
avg (Imp_Abril)	2,86	3,44	4,23	2,52
avg (Frec_Abril)	2,93	3,31	4,33	2,43
avg (Imp_Mayo)	2,92	3,57	4,28	2,46
avg (Frec Mayo)	3,00	3,36	4,32	2,38
avg (Imp_Junio)	2,94	3,68	4,15	2,44
avg (Frec_Junio)	3,05	3,32	4,29	2,40
avg (Imp_Julio)	2,93	3,42	4,21	2,51
avg (Frec_Julio)	2,99	3,26	4,34	2,40
avg (Imp_Agosto)	2,88	3,31	4,28	2,60
avg (Frec_Agosto)	2,94	3,43	4,27	2,44
avg (Imp_Septiembr...)	2,88	3,74	4,21	2,39
avg (Frec_Septiembr...)	2,96	3,45	4,38	2,38
avg (Imp_Octubre)	2,92	3,74	4,24	2,34
avg (Frec_Octubre)	3,00	3,46	4,26	2,39
avg (Imp_Noviembr...)	2,87	3,73	4,17	2,44
avg (Frec_Noviembr...)	2,97	3,47	4,25	2,41
avg (Imp_Diciembre...)	2,88	3,46	4,22	2,56
avg (Frec_Diciembre...)	2,89	3,48	4,31	2,39

Cluster	1	2	3	4
avg (Imp_Lunes)	2,91	3,61	4,11	2,54
avg (Frec_Lunes)	2,99	3,43	4,29	2,37
avg (Impn Martes)	2,89	3,64	4,18	2,46
avg (Frec Martes)	3,04	3,34	4,28	2,40
avg (Imp_Miercoles)	2,91	3,44	4,19	2,53
avg (Frec_Miercoles)	3,05	3,34	4,25	2,43
avg (Impn Jueves)	2,92	3,66	4,15	2,48
avg (Frec Jueves)	3,00	3,46	4,24	2,42
avg (Impn Viernes)	2,83	3,72	4,13	2,52
avg (Frec Viernes)	2,93	3,50	4,28	2,39
avg (Impn Sabado)	2,80	3,33	4,42	2,40
avg (Frec Sabado)	2,86	3,30	4,36	2,44
avg (Imp_Domingo)	2,92	3,20	4,31	2,50
avg (Frec_Domingo)	2,90	3,33	4,32	2,45
avg (Imp_Manana)	2,87	3,81	3,98	2,70
avg (Frec_Manana)	2,82	3,37	4,29	2,51
avg (Imp_Tarde)	2,93	3,33	4,38	2,37
avg (Frec_Tarde)	3,05	3,40	4,29	2,36
avg (Imp_Noche)	2,89	3,34	4,15	2,63
avg (Frec_Noche)	2,93	3,33	4,16	2,61

Tabla 46: Medias de cada variable en cada clúster. Método fuzzy

4.2.6. Comparativa de resultados de los diferentes análisis clúster

Comparativa k-means frente a jerárquico

- De los 24 consejeros del clúster 1 del k-means, 18 de ellos forman parte del clúster 3 jerárquico, lo que supone el 78% del clúster jerárquico total. Ambos clúster cuentan como características comunes ser los que más gastan en Ocio, los que más gastan y con mayor frecuencia en telefonía y moda, los que la emplearon en el supermercado con mayor frecuencia.

De los 6 consejeros restantes del clúster del k-means, 4 pertenecen al cluster2 jerárquico y 2 al clúster 4.

- De los 10 consejeros del clúster 2 del k-means, 5 de ellos están clasificados dentro del clúster 3 jerárquico y los otros 5 están dentro del clúster 4. En el caso del k-means, se trata de los consejeros que mayor gasto hicieron en casi todas las variables y en el caso de los clúster 3 y 4 se tratan de los consejeros que hicieron también mayor gasto pero discriminando en algunas variables concretas (por ejemplo en el clúster 3 fue en comercios, telefonía, restaurantes, supermercados, moda, gasolineras y ocio, y en el clúster 4 fue en hoteles, viajes y cajeros), siendo por tanto el clúster jerárquico en este caso más explicativo.
- De los 18 consejeros del clúster 3 del k-means, 14 pertenecen al clúster 2 del jerárquico, suponiendo el 67% del clúster jerárquico total. Ambos clúster cuentan como característica común el ser los que más gastaron durante el año 2012. Los otros 4 consejeros pertenecen al clúster 1.
- De los 29 consejeros del clúster 4 del k-means, 28 pertenecen al clúster 1 del jerárquico, suponiendo el 87% del clúster jerárquico total. Ambos clúster cuentan como característica común ser los que menor gasto y menor frecuencia de uso hicieron de las tarjetas en casi todas las variables de estudio. El otro consejero pertenece al clúster 2.

Comparativa k-means frente a fuzzy

- De los 24 consejeros del clúster 1 del k-means, 18 pertenecen al clúster 3 del fuzzy, suponiendo el 64,3% del clúster fuzzy total. La característica común entre estos clusters es que nos clasifican a los consejeros con mayor gasto y frecuencia de uso; sin embargo el clúster k-means es más específico indicando las variables en que mayor fue el gasto (ocio, moda, supermercados, telefonía) y el clúster fuzzy es más genérico indicando que el gasto y frecuencia fueron mayores en todas las variables en general.

Otros 6 de los 24 consejeros pertenecen al clúster fuzzy 2, suponiendo el 50% del clúster fuzzy. La única característica común entre estos clúster es que en ambos la frecuencia de uso de las tarjetas durante el 2012 fue la mayor.

- De los 10 consejeros del clúster 2 del k-means, 9 pertenecen al clúster 3 del fuzzy, suponiendo el 32% del clúster fuzzy total. La característica común entre estos dos clusters es que ambos recogen a los consejeros que mayor gasto y frecuencia de uso de las tarjetas hicieron.

El consejero restante pertenece al clúster 2 del fuzzy.

- De los 18 consejeros del clúster 3 del k-means, 13 pertenecen al clúster 1 del fuzzy, suponiendo el 81,3% del clúster fuzzy total. La característica común entre estos clúster es que en ambos, los consejeros hicieron el mayor gasto durante el año 2012.

Los otros 5 consejeros pertenecen al clúster 2 del fuzzy, suponiendo el 42% del clúster fuzzy total. La característica común entre estos clúster es que en ambos, los consejeros hicieron el mayor uso de las tarjetas durante el año 2012.

- De los 29 consejeros del clúster 4 del k-means, 26 pertenecen al clúster 4 fuzzy, suponiendo el 100% del clúster fuzzy. La característica común entre estos dos clusters es que ambos recogen a los consejeros que menor gasto y frecuencia de uso de las tarjetas hicieron en casi todas las variables.

Los otros 3 consejeros pertenecen al clúster 1 fuzzy, suponiendo el 18,75% del clúster.

Comparativa jerárquico frente a fuzzy

- De los 31 consejeros del clúster 1 del jerárquico, 24 pertenecen al clúster 4 fuzzy, suponiendo el 92,3% del clúster fuzzy. La característica común entre estos dos clusters es que ambos recogen a los consejeros que menor gasto y frecuencia de uso de las tarjetas hicieron en casi todas las variables.

Los otros 7 consejeros del clúster jerárquico pertenecen al clúster 1 fuzzy, suponiendo el 43,75% del clúster fuzzy. En este caso, estos dos grupos únicamente coinciden en tener los consejeros con menor gasto durante el año 2003.

- De los 20 consejeros del clúster 2 jerárquico, 8 pertenecen al clúster 2 fuzzy, lo que supone el 66,7% del clúster fuzzy. La característica común entre estos dos clusters es que contienen los consejeros con mayor frecuencia de uso durante el año 2012.

Otros 9 consejeros del clúster jerárquico pertenecen al clúster 1 fuzzy, lo que supone el 56,25% del clúster fuzzy. La característica común entre estos dos clusters es que contienen los consejeros con mayor gasto durante el año 2012.

Los dos consejeros restantes pertenecen al clúster 4 fuzzy, pero en este caso no se encuentran similitudes entre ambos clusters.

El último consejero (Vela García, Carlos) pertenece al clúster 3 fuzzy, y en este caso tampoco se encuentran similitudes entre ambos clusters.

- De los 23 consejeros del clúster 3 jerárquico, 21 pertenecen al clúster 3 fuzzy, lo que supone el 75% del clúster fuzzy. La característica común entre estos dos clusters es que contienen los consejeros con mayor gasto en comercios, telefonía, restaurantes, supermercados, moda, gasolineras y ocio.

Los otros 2 consejeros pertenecen al clúster 2 fuzzy, lo que supone el 16,7% del clúster fuzzy. La característica común entre estos dos clusters es que contienen los consejeros con mayor frecuencia de uso durante el año 2012.

- De los 7 consejeros del clúster 4 jerárquico, 5 pertenecen al clúster 3 fuzzy, lo que supone el 17,85% del clúster fuzzy. La característica común entre estos dos clusters es que fueron los consejeros que más gastaron en hoteles, viajes y cajeros y entre 2004 y 2010. Se podría considerar el clúster jerárquico como un subgrupo del clúster fuzzy.



Los otros 2 consejeros pertenecen al clúster 2 fuzzy, pero en este caso no se encuentran similitudes entre ambos clusters.

Tras esta comparativa podemos ver cómo tanto el clúster jerárquico como el k-means nos ofrecen unos clusters con mayor información y más interpretable, considerando por tanto las conclusiones obtenidas con el clúster fuzzy como de peor calidad, ofreciéndonos menos detalla y explicabilidad de los clúster.

Tanto el clúster jerárquico como el k-means nos proporcionan mayor información de los consejeros, pero podemos considerar como mejor método el análisis jerárquico ya que nos ha proporcionado clusters con mayor información de los consejeros.

5.CONCLUSIONES

Como datos más significativos de todo el estudio del gasto realizado por los consejeros con las tarjetas Black hemos obtenido las siguientes conclusiones:

- Los consejeros Enrique de la Torre Martínez, Ricardo Morado Iglesias, José Antonio Moral Santín, Ramón Ferraz Ricarte y Domingo Naval Moral Sánchez fueron los que mayor gasto realizaron en cajeros y con las tarjetas.
- Los consejeros Carlos Vela García, Carmen Contreras Gómez, Rafael Eduardo Torres Posada, María Carmen Cafranga Cavestany, Luis Gabarda Durán, Juan Emilio Iranzo Martín y Ramón Martínez Vilches son los que más gastaron en viajes y moda.
- Los consejeros Jorge Gómez Moreno, Gonzalo Martín Pascual y José María Fernández del Río Fernández fueron los que más gastaron en restaurantes.
- La mayor parte de los consejeros centran el gasto realizado en la operación “Compra”
- El consejero Ricardo Morado Iglesias se caracteriza por ser el que mayor gasto hizo en “Disposición de efectivo en Oficinas”.
- El consejero José Antonio Moral Santín se caracteriza por ser el que mayor gasto hizo en “Reintegro en cajero Propio”.
- Los consejeros José Ricardo Martínez Castro, Ramón Martínez Vilches y José María Arteta Vico fueron los que mayor gasto realizaron por la noche.
- La mayor parte del gasto se realizó en el tramo de la tarde.
- Los consejeros José Caballero Domínguez y Miguel Corsini Freese son los que más gastaron por la mañana.
- Los consejeros José María Arteta Vico, Rodrigo de Rato Figaredo y Pedro Bedia Pérez fueron los que mayor gasto realizaron en Domingo.
- De los tres análisis clúster realizados el análisis fuzzy es el que menor información nos proporciona, quedándonos por tanto con la información aportada bien por el análisis jerárquico o el análisis k-means.
- En el análisis jerárquico, se ha podido segmentar a los consejeros en 4 categorías diferentes:
 - Los que menor gasto realizaron y con menor frecuencia emplearon las tarjetas
 - Los que mayor número de veces y más gasto realizaron en el año 2012
 - Los que mayor gasto y con mayor frecuencia usaron las tarjetas en comercios, telefonía, restaurantes, supermercados, moda, gasolineras y ocio. Además, en la mayor parte de los años eran los que más frecuentemente empleaban la tarjeta pero no los que más gasto hacían con ella.

- Los que mayor gasto hicieron en hoteles, viajes y cajeros. Entre 2004 y 2010 fueron los que más gastaron y los que menos entre 2011 y 2012. Sin embargo, en los años de mayor gasto no fueron los que más veces emplearon las tarjetas, por lo que las compras que realizaban eran de mayor importe que la media del resto de consejeros.
- En el análisis k-means, se ha podido segmentar a los consejeros en 4 categorías diferentes:
 - Los que más viajan, más gastan en ocio, más y con más frecuencia en moda, supermercados y telefonía.
 - Los que más gastaron y con mayor frecuencia
 - Los que pasaron de hacer el menor uso de la tarjeta y menor gasto durante el 2011 a ser los de mayor gasto y uso durante el 2012.
 - Los que menos gastaron y con menor frecuencia.
- En la segmentación realizada se ha visto también cómo el consejero Ildefonso José Sánchez Barcoj empleaba la tarjeta y gastaba con ella cantidades muy superiores a la media de los otros consejeros.

Como se indicó en el punto 1.2. del presente documento, el interés de este estudio radica por una parte en el posible uso fraudulento de las tarjetas por parte de los consejeros, ya que eran conscientes de que no declaraban a Hacienda el gasto que hacían con ellas ni tenían necesidad de justificarlo pero aún así hicieron uso de las mismas incluso para propósitos para los que inicialmente no estaban destinadas. Por otra parte, el dinero gastado con esas tarjetas provenía de un banco rescatado con dinero público, lo que agrava por tanto más la situación al ser dinero puesto por todos los contribuyentes.

Tras el estudio realizado y a la vista de los resultados obtenidos, aun no siendo especialmente comprometedores, hace pensar en la posible necesidad de realizar algún estudio o auditoría sobre las “retribuciones” de este tipo contempladas por otros bancos españoles.

6. BIBLIOGRAFIA

ABC [2015]. Bankia, el rescate más caro en la historia de España 23.465 millones de euros. [sitio web]. [Consulta: 28/01/2015]. Disponible en:

<http://www.abc.es/20120526/economia/abci-bankia-nacionalizacion-ayudas-publicas-201205252315.html>.

ANALISIS Y DECISION [2015]. *Manual. Curso introducción de R. Capítulo 17: Análisis Clúster con R (y III)*. [sitio web]. [Consulta: 10/09/2015]. Disponible en:

<http://analisisydecision.es/manual-curso-introduccion-de-r-capitulo-17-analisis-cluster-con-r-y-iii/>

CINCO DIAS [2015]. Consulta la lista de las tarjetas 'B' de Caja Madrid y cuánto se gastaron sus 'dueños'. [sitio web]. [Consulta: 26/01/2015]. Disponible en:

http://cincodias.com/cincodias/2014/10/03/mercados/1412335445_083103.html.

CINCO DIAS [2015]. Diez claves sobre las tarjetas opacas de Caja Madrid: qué son, quiénes las recibieron y por qué se investigan. [sitio web]. [Consulta: 26/01/2015]. Disponible en:

http://cincodias.com/cincodias/2014/10/11/mercados/1413035129_754098.html.

DÍAZ COVIÁN, IÑAKI [2014]. Descubrimiento de conocimiento en bases de datos espaciales. [Sitio web]. [Consulta: 12/10/2015]. Disponible en:

http://dspace.sheol.uniovi.es/dspace/bitstream/10651/27914/6/TFM_I%C3%B1akiDiazCovian.pdf

DOMINO [2015]. *Visualizing home ownership with small multiples and R*. [sitio web]. [Consulta: 02/09/2015]. Disponible en:

<http://blog.dominodatalab.com/visualizing-homeownership-in-the-us-using-small-multiples-and-r/>

EL DIARIO [2015]. Anatomía del rescate a Bankia: los últimos días de las tarjetas 'Black'. [sitio web]. [Consulta: 28/01/2015]. Disponible en:

http://www.eldiario.es/politica/Bankia_0_312169537.html.

EL DIARIO [2015]. *Diez preguntas sobre las tarjetas 'Black' de Caja Madrid*. [sitio web]. [Consulta: 26/01/2015]. Disponible en:

http://www.eldiario.es/escolar/preguntas-tarjetas-black-Caja-Madrid_6_310478966.html

EL DIARIO [2015]. Los consejeros de Caja Madrid tenían tarjetas de crédito en 'negro' de hasta 50.000 euros al año. [sitio web]. [Consulta: 26/01/2015]. Disponible en:

http://www.eldiario.es/economia/consejeros-Caja-Madrid-tarjetas-limites_0_206830123.html.



USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

EL PAIS [2015]. *Imputación general por las tarjetas fantasma de Caja Madrid*. [Sitio web]. [Consulta: 28/01/2015]. Disponible en:

http://politica.elpais.com/politica/2015/01/28/actualidad/1422441939_245671.html.

ETH ZÜRICH UNIVERSITY [2015]. *Fuzzy Analysis Clustering*. [sitio web]. [Consulta: 10/10/2015]. Disponible en: <https://stat.ethz.ch/R-manual/R-devel/library/cluster/html/fanny.html>

ETH ZÜRICH UNIVERSITY [2015]. *K-Means Clustering*. [sitio web]. [Consulta: 10/10/2015]. Disponible en: <https://stat.ethz.ch/R-manual/R-patched/library/stats/html/kmeans.html>

ETH ZÜRICH UNIVERSITY [2015]. *Partitioning Around Medoids*. [sitio web]. [Consulta: 02/10/2015]. Disponible en:

<https://stat.ethz.ch/R-manual/R-devel/library/cluster/html/pam.html>

FRANCESC CARMONA [2014]. Un ejemplo de ACP paso a paso. Apuntes, Departamento de estadística. Universidad de Barcelona. [sitio web]. [Consulta: 23/10/2015].

<http://www.ub.edu/stat/docencia/Mates/ejemploACP.PDF>

GEOCITIES [2015]. Curso de R: Capítulo 10: Análisis de conglomerados (clúster) I. [sitio web]. [Consulta: 04/10/2015]. http://www.geocities.ws/r_vaquerizo/Manual_R10.html.

GRANÉ, AUREA [2015]. Distancias estadísticas y Escalado Multidimensional (Análisis de Coordenadas Principales). Apuntes, Departamento de Estadística. Universidad Carlos III de Madrid.

HUFFINGTON POST [2015]. *Los cuatro 'quiénes' del caso de las tarjetas de Caja Madrid* [sitio web]. [Consulta: 26/01/2015]. Disponible en:

http://www.huffingtonpost.es/pablo-herreros-lavina/los-cuatro-quienes-del_b_5940800.html.

INVERSION & FINANZAS [2015]. *Historia del florecimiento y declive de Bankia* [sitio web]. [Consulta: 26/01/2015]. Disponible en:

<http://www.finanzas.com/noticias/mercados/bolsas/20120719/historia-florecimiento-declive-bankia-1464209.html>.

JOHNSON, DALLAS E. 1938. *Métodos multivariados aplicados al análisis de datos*. México: International Thomson. ISBN: 968-7529-90-3

LA INFORMACION [2015]. Las tarjetas 'Black' de Caja Madrid: historia de los seres insaciables. [sitio web]. [Consulta: 27/01/2015]. Disponible en:

<http://blogs.lainformacion.com/zoomboomcrash/2014/10/05/las-tarjetas-black-de-caja-madrid-historia-de-los-seres-insaciables/>.



USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

LIBRARY OF THE COLLECTIVE HUMAN RECORD [2015]. Curso de R: Capítulo 9: Análisis de Componentes Principales. [Sitio web]. [Consulta: 04/10/2015]. Disponible en:

http://www.lchr.org/a/55/gy/Manual_R9.htm

SHARP SIGHT LABS [2014]. How to make a small multiples chart in R. [sitio web]. [Consulta: 14/10/2015]. Disponible en:

<http://www.sharpsightlabs.com/small-multiples-ggplot/>

SUSI GARCIA, ROSARIO. 2014. *Metodología de un Trabajo de Investigación*. Apuntes, Facultad de Estudios Estadísticos. Universidad Complutense de Madrid

TARINGA [2015]. *Las tarjetas de crédito más exclusivas del mundo*. [sitio web]. [Consulta: 26/01/2015]. Disponible en:

<http://www.taringa.net/posts/imagenes/16487553/Las-tarjetas-de-credito-mas-exclusivas-del-mundo.html>.

20 MINUTOS [2015]. El juez imputa a 78 consejeros y directivos de Caja Madrid y Bankia por las 'tarjetas Black'. [sitio web]. [Consulta: 28/01/2015]. Disponible en:

<http://www.20minutos.es/noticia/2359910/0/imputados/consejeros-directivos/tarjetas-black/>.

20 MINUTOS [2015]. Cinco verdades incómodas sobre las tarjetas “fantasma” de Caja Madrid/Bankia. [sitio web]. [Consulta: 28/01/2015]. Disponible en:

<http://blogs.20minutos.es/un-cuento-corriente/2014/10/03/cinco-verdades-incomodas-sobre-las-tarjetas-fantasma-de-caja-madridbankia/>.

UNIVERSIDAD COMPLUTENSE DE MADRID [2015]. *Clasificación y ordenación con r*. [sitio web]. [Consulta: 08/09/2015]. Disponible en:

www.um.es/docencia/geobotanica/ficheros/practica2.pdf

UNIVERSIDAD DE GRANADA [2015]. Ejemplos de análisis clúster. [sitio web]. [Consulta: 04/10/2015]. Disponible en:

<http://www.ugr.es/~mvargas/3.DosEjesanalisisclusteryCCAA.pdf>

UNIVERSIDAD DE GRANADA [2015]. Ejemplo de análisis de componentes principales. [sitio web]. [Consulta: 23/10/2015]. Disponible en:

<http://www.ugr.es/~mvargas/4.Ejemplo1ACP.pdf>

VALENCIA DELFA, JOSE LUIS. 2013. *Complementos de Formación en Técnicas de Minería de Datos*. Apuntes, Facultad de Estudios Estadísticos. Universidad Complutense de Madrid



USO Y ABUSO DE LAS TARJETAS BLACK POR CONSEJEROS DE BANKIA

VALENCIA DELFA, JOSE LUIS. 2013. *Técnicas y Metodología de la Minería de Datos (SEMMA)*. Apuntes, Facultad de Estudios Estadísticos. Universidad Complutense de Madrid

VOXPOPULI [2015]. Las tarjetas opacas de Caja Madrid se otorgaban "por sensibilidades políticas": más a los más afines. [sitio web]. [Consulta: 26/01/2015]. Disponible en:

<http://vozpopuli.com/actualidad/50423-las-tarjetas-opacas-de-caja-madrid-se-otorgaban-por-sensibilidades-politicas-mas-a-los-mas-afines>.

WIKIPEDIA [2015]. *Bankia*. [sitio web]. [Consulta: 28/01/2015]. Disponible en:

http://es.wikipedia.org/wiki/Bankia#Rescate_financiero_y_nacionalizaci.C3.B3n.

YANCHANG ZHAO. 2013. *R and Datamining: Examples and case studies*. Australia: Academic Press, Elsevier. ISBN: 978-0-123-96963-7